Drug Evaluation From Now On

What's the value of local data in the global clinical data package?

FY2020 Task Force 5,

Data Science Expert Committee,

Drug Evaluation Committee,

Japan Pharmaceutical Manufacturers Association

July 2022

English version issued in December, 2022

Table of Contents

Columns

Introduction

Global simultaneous drug development is a norm. Global development has become an inevitable development strategy not only for global companies that are considering the simultaneous introduction of drugs to the global market from the beginning, but also for Japanese companies that are pursuing a strategy of expanding their markets worldwide while developing in Japan first. This way of thinking is not limited to Japan, but would be also true for other countries that are actively participating in global development (although each country has its own history and environmental background). The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) was established in 1990 with the aim of harmonizing regulations among the three regions of Japan, the United States, and Europe. In December 2021, the ICH published the "Overview of ICH," [1] which summarizes the purpose of the ICH as follows.

Purpose of ICH

Promotion of public health through **international harmonization** that contributes to

- Prevention of unnecessary duplication of clinical trials and postmarket clinical evaluations
- Development and manufacturing of new medicines
- Registration and supervision of new medicines
- Reduction of unnecessary animal testing without compromising safety and effectiveness

Accomplished through **Technical Guidelines** that are implemented by the regulatory authorities

Directly related to this report is the first bullet "Preventing unnecessary duplication of clinical trials and post-marketing clinical evaluations," which states that drug evaluations should be considered from a global perspective and duplication of clinical trials and postmarket clinical evaluations should be avoided. The purpose of ICH has not changed from the original one. However, as the number of regulators and organizations participating in ICH has increased

[1] ICH, Overview of ICH, December 2021 (see p. 4), (https://admin.ich.org/sites/default/files/2021-12/OverviewOfICH_2021_1202.pdf; last access confirmed May 25, 2022) May 25, 2022)

dramatically[2], it is important to reiterate that ICH is no longer a trilateral regulatory harmonization meeting among Japan, the US, and Europe, and that all parties involved in drug evaluation, including regulators, industry, academia, and patients from around the world, share the objectives of the ICH. Under these circumstances, don't we need to change our thought patterns and strategies regarding drug evaluation and our actions based on these (reflected in documents for the communication to regulatory authorities)? The FY2020 Task Force 5 of the Data Science Expert Committee, the Drug Evaluation Committee, the Japan Pharmaceutical Manufacturers Association (JPMA) considered these issues as a starting point for its discussions, which led to the compilation of this report.

The drug evaluation leading up to regulatory submission is a learning process to determine from a global perspective whether the drug is effective and safe for patients with various backgrounds (patient backgrounds and environments) and whether the balance of benefits and risks is acceptable. On the other hand, in each country (for us, Japan), how to estimate the efficacy and safety of the drug for patients who will use the drug in the future in that country. This is an important basis for regulatory authorities to make decisions on the approval of pharmaceutical products. For example, in Japan, data collected in Japan have traditionally been used as the core data, and comparisons have been made between Japan and other countries (or overall study population including Japan) to show similarities and consistency. If Japanese data shows different responses, the reasons for the difference are discussed. If such an approach is acceptable, a similar approach may be justified and repeated for any country participating in global development. Furthermore, there may be confusion arising from the different evaluation policies of the various regulatory authorities and from each country's insistence on securing its regional sample size. If such chaos were to spread, it would clearly contradict the objectives of the ICH described above, and would be scientifically questionable.

The approval of a drug is a major milestone in the learning process for the efficacy and safety of a given drug, but it is not the goal. The learning process continues even after the drug is marketed. While postmarketing drug evaluation has the aspect of advancing the drug to be safer, more effective, and easier to use over the world, it also has the aspect of evaluating the efficacy and safety of a drug in the local environment, including the medical environment, as well as the value of the drug to the society. We believe that deepening our understanding of factors affecting the treatment effect (effect modifiers) will provide a new axis for thinking

[2] As of June 2022, the number of member and observer regulators and organizations will exceed 50. (https://www.ich.org/page/members-observers; last access confirmed June 9, 2021)

and lead to a change in thinking. We are now convinced that thinking centered on effect modifiers is scientific thinking that can be applied not only to global development, but also to drug development under various other circumstances. Although the discussion in this report will focus on global development, it will also include thinking that can be applied to other drug development when evidence, such as foreign data and literature information, are available. Let us begin with scientific thinking, which is the home to which we should return whenever facing difficulties.

Fundamentals for Improving Evidence

Throughout the life cycle of a drug, from development to postmarketing, there are various **learning processes to** understand the efficacy and safety of the drug. In the learning process, various types of studies and trials, including nonclinical studies and clinical studies/research, can be conducted continuously to increase the evidence regarding important findings in drug evaluation and to consolidate these findings.

The learning process in science, including drug evaluation, is an iterative process **of deduction** and **induction of** various propositions (Note), through which knowledge is continually updated, information uncertainty is reduced, and confidence is deepened.

(Note: A proposition is a statement of the content of a single judgment in logic, such as "A is B." In mathematics, it is something that is determined in principle to be true or false in its theory.[3] Examples of propositions in drug evaluation are:

- For an endpoint X, drug A is more effective than control drug B
- Factor X is an effect modifier that has a significant effect on the therapeutic effect of drug A
- Drug A causes adverse drug reaction B

There are two levels of **propositions**: the level of "what can be said **in general (GENERAL)"** as principles that can be applied in many cases, and the level of **"**what can be said **in particular situations (PARTICULAR)**" (Figure 1). Thinking that tries to apply from General to Particular is called **deduction**, and thinking that tries to generalize from Particular to General is called **induction**. PARTICULAR is what we can observe and experience, and GENERAL is a generalized concept or principle. Examples of PARTICULAR are the results of a clinical trial or a portion of it, i.e., a finding from a subgroup analysis, or a variety of **findings** from exploratory analyses. We try to conceptualize the findings in PARTICULAR to the level of GENERAL, which is generally considered valid, for future use as experience. The next time we experience a PARTICULAR, we consider whether the GENERAL supported by our experience

[3] Shinmeikai Japanese Dictionary, 7th edition, Sanseido (2011)

can be applied to the PARTICULAR in front of us. If it is successfully applied, the GENERAL will be further solidified, and if not, the GENERAL will be updated. The learning process of induction and deduction continues like this.[4]

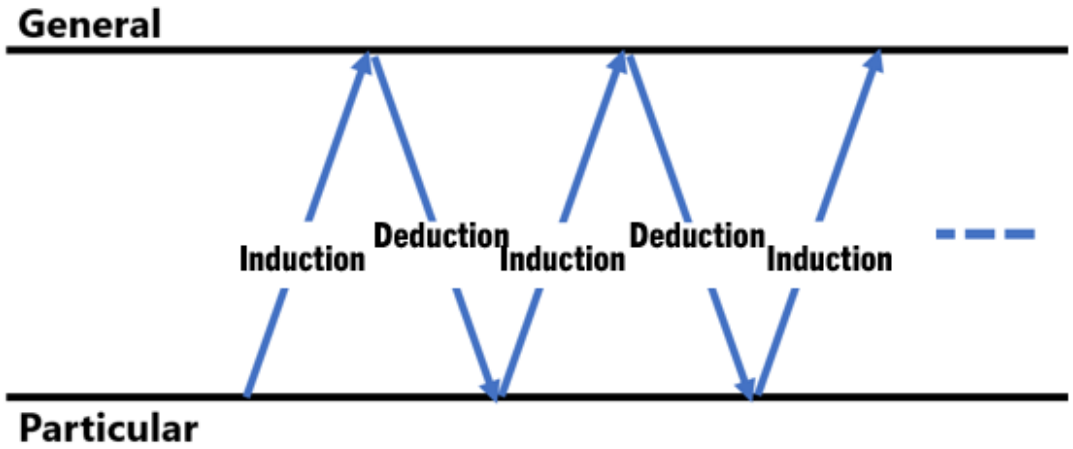


Figure 1 Scientific learning process: repetition of induction and deduction

This learning process is broadly consistent with our thought patterns. For example, a young person newly entering the company experiences PARTICULAR in his/her respective workplace and area of expertise. He/she realizes that he/she can generalize and conceptualize his/her own experiences in his/her own thinking and forms a GENERAL in his/her mind. Next, he/she tries to deduce if he/she can make use of his/her past experience in the given job. If it works, the GENERAL he/she had is reinforced. If not, he/she will think about why the GENERAL he/she had could not be applied, and how he/she can update it to encompass this PARTICULAR. Over years of this kind of thinking, he/she become skilled at the work.

The learning process involved in the efficacy and safety of drugs is essentially the same as this thought pattern. Figure 2 is a modified version of a diagram originally drawn by George Box, one of the leading applied statisticians of the 20th century, in his book.[5] There are other explanations of increasing knowledge through the scientific learning process, a prominent one being Kolb's Experimental Learning Theory.[6] In Figure 2, the scientific learning process of induction and deduction is depicted in the form of a loop, and the explanation focuses on the updating of GENERAL.

---

[4] Bosland W.M., Curran, J.M. Introduction to Bayesian Statistics, 3rd Edition (Wiley ,2016)

[5] Box, E.P.G., J., Hunter, J.S., Hunter, W.G., Statistics for Experimenters: Design, Innovation, and Discovery, 2nd Edition (Wiley, 2005)

[6] Experiential learning: helpful review of sites by Tim Pickles (https://reviewing.co.uk/research/experiential.learning.htm; last accessed 22 May 2021 May 22, 2021)

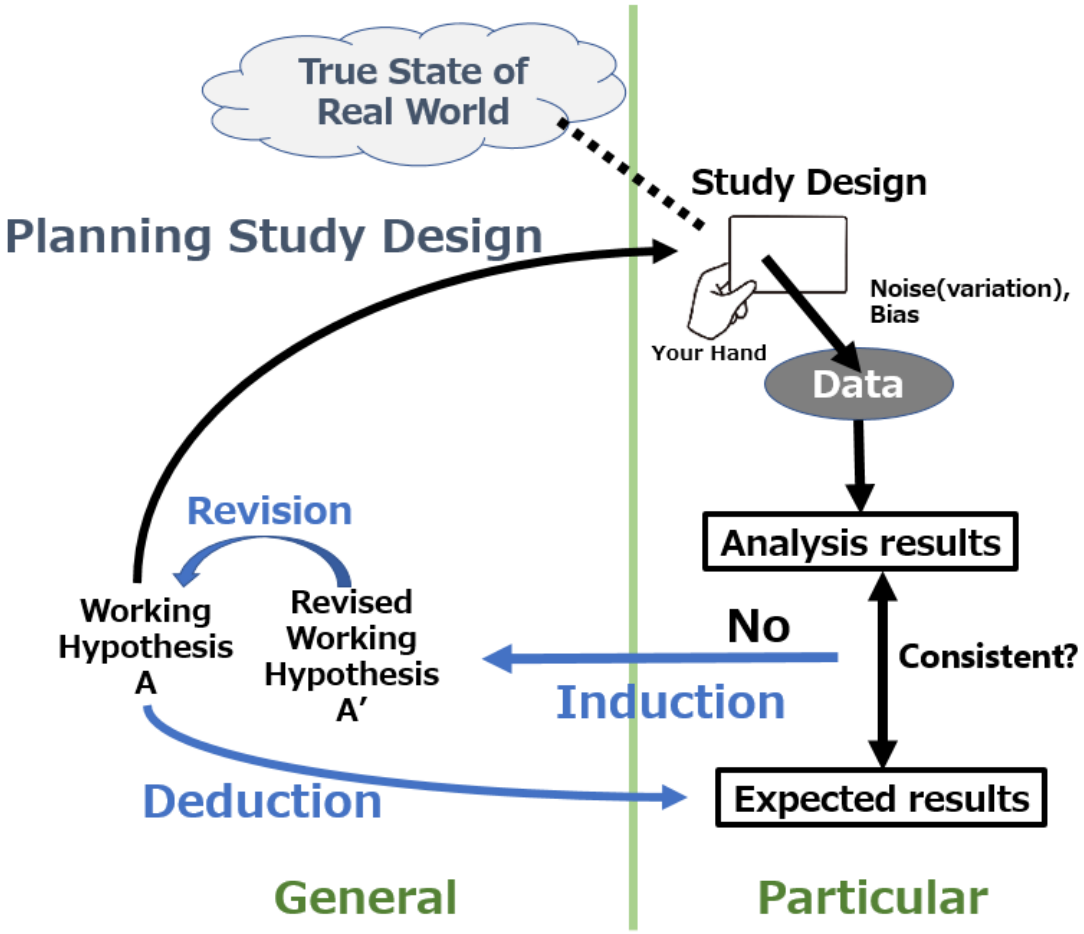


Figure 2 Scientific learning process

Your first GENERAL or working hypothesis (e.g., "The drug is found to be effective in the primary endpoint X") is on the left side of Figure 2 as "Working Hypothesis A." You deductively ask, "If A is correct, what results would be observed?" The arrow extending from A to the lower right is the result predicted by the deduction. The arrow from A to the upper right depicts the study you design as a **square window.** The hand holding the window is your hand, and what kind of window you design is in your hands. Through this window, you will see an aspect of the real world as data, but the data may be clouded by noises or biased. Your analysis of the data from the study you designed will be compared to the expected results. Both the expected results and the results derived from the data are at the level of PARTICULAR. If they agree, the proposition is proven, and the loop ends. This is the case when the primary efficacy endpoint is achieved in a confirmatory study. The expected results are prespecified and clearly stated in the protocol and/or statistical analysis plan of the study. When the results of the data analysis are as expected, the proposition (hypothesis directly

related to the purpose of the study) has been confirmed and that provides a high level of evidence to support that the proposition is true. If the results of the data analysis do not match the expected results, the learning process may be terminated (development stopped). If the learning process continues, then working hypothesis A is updated and a new working hypothesis A' is established. This is induction. A new study is conducted based on the updated GENERAL, working hypothesis A', and the loop continues.

If the confirmation is explicit, as in the case of the primary efficacy endpoint, the explanation in Figure 2 would be easy to understand. However, in addition to confirmatory propositions, there are various other important propositions in the life cycle of a drug that relate to its efficacy and safety. The following propositions are examples.

[Important propositions in drug evaluation, not always evaluated in the framework of confirmation].

- Propositions other than confirming propositions about the primary endpoints
- Propositions related to secondary endpoints not included in the confirmation framework
- Propositions in subgroups defined by some attributes
- Propositions on effect modifiers
- Propositions related to safety assessment not intended for confirmation

These are propositions in which the learning process proceeds concurrently during the life cycle of a drug, and we can say that the majority of our drug evaluations, with a goal to "***Know the drug,*"** are such propositions. The learning process for these propositions should also involve the repetition of deduction and induction as shown in Figure 1, but the learning process often does not proceed in the form shown in Figure 2. Then, how should we repeat the process of deduction and induction?

It is not likely that a single trial or study will provide sufficient evidence for a given proposition. That is why it is necessary to continuously evaluate the information and study results obtained during the drug evaluation process. In order to accumulate findings that, standing alone, do not provide a sufficient explanation, and to increase the evidence in support of the proposition's truth, the following thinking is important[7].

[7] Drug Safety Data: How to Analyze, Summarize, and Interpret to Determine Risk, translated into Japanese by Osamu Komiyama, Hironori Sakai, Tomomi Kimura, et al. (Scientist, 2012) originally published in Klepper, M.J., Cobert, B., (Jones & Bartlett Learning, 2011))

The more findings that point in the same direction (support the same conclusion), the more confident we are that they are true signals.

The perspectives for evaluating efficacy and safety based on this concept are described as **internal consistency** and **external consistency** in the text of the E17 Guideline and in Training Material Module 6.[8] These two perspectives, plus biological rationale, clinical significance, and statistical uncertainty, are summarized as **five perspectives** (Figure 3).

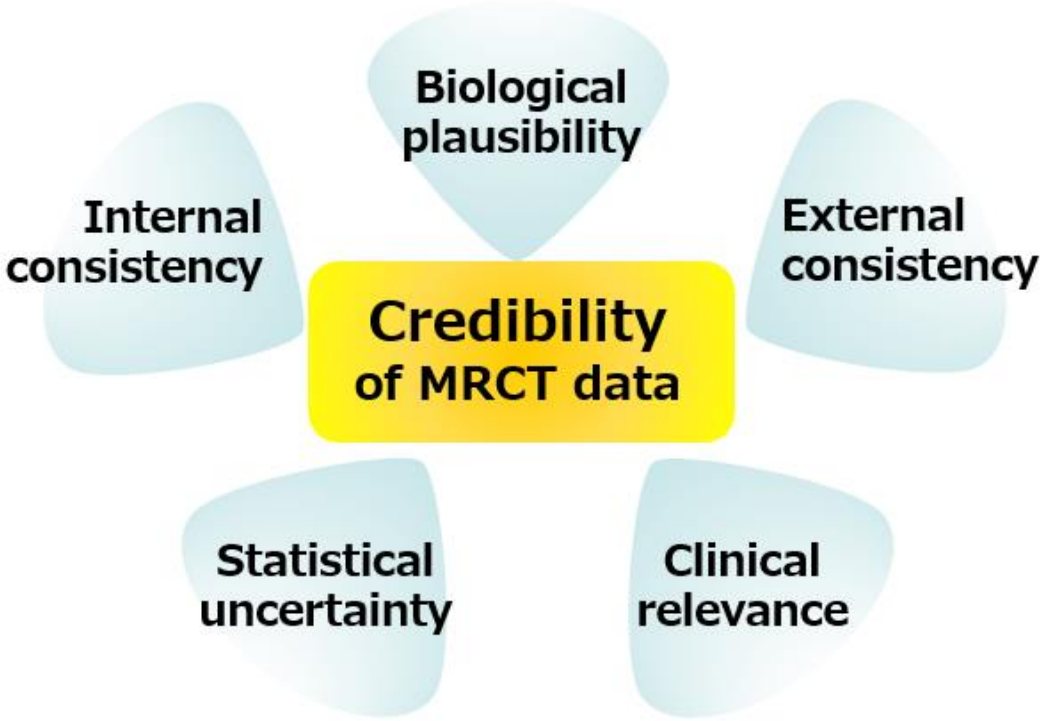


Figure 3 Five perspectives for assessing the credibility of MRCT results; pages 8-9 of ICH E17 Training Material Module 6

Figure 3 summarizes the perspectives for evaluating the credibility of the results of a **multi-regional clinical trial (MRCT)**. They are widely applicable not only to MRCTs but also to the

[8] Training Materials on "General Principles for the Planning and Design of Global Clinical Trials" (https://www.pmda.go.jp/int-activities/int-harmony/ich/0022.html; last access confirmed May 20, 2022)

evaluation of evidence for [Important propositions in drug evaluation, not always evaluated in the framework of confirmation].

Biological persuasiveness:	Is a convincing biological explanation possible?
Clinical Relevance:	Are the findings important in providing a basis for clinical judgment or treatment decisions? Are the differences observed among any populations clinically noteworthy?
Statistical Uncertainty:	How certain are the findings obtained? In other words, is there any non-negligible bias in the estimation, and how accurate is the estimation?
Internal Consistency:	Multiple findings supporting each other *within* the same study, e.g., findings supporting the same conclusion (proposition) on multiple biologically or medically relevant endpoints.
External Consistency:	When the findings from one study are viewed side-by-side with external information, such as the results of other studies, do the findings support the same conclusion (proposition)?

The evaluation of evidence for [Important propositions in drug evaluation, not always evaluated in the framework of confirmation] should be conducted from these five perspectives. The individual findings are evaluated in terms of their clinical significance and weight as evidence in support of a proposition, taking into account statistical uncertainty. Findings from a single trial are further enhanced by checking their relevance and whether they support the same conclusion (internal consistency), and by checking their consistency and coherence with other trials and known information (external consistency). If, in the process of checking for internal and external consistency, each finding is also biologically reasonable, the belief that the proposition is true will be further reinforced.

What are Effect Modifiers?

The response to a drug varies from person to person. Some drugs work well, while others do not respond as well as expected. Some people experience adverse drug reactions that interfere with their lives or prevent them from continuing treatment, while others experience few adverse events. These differences in drug response arise from individual background characteristics such as demographic attributes, genetic characteristics, medical history, and health status, as well as living and medical environments.[9] In general, individual background characteristics can be categorized as **patient factors**, while living and medical environments can be categorized as **environmental factors**. In the context to explain ethnic differences in the global development of drugs, the former is classified as **an intrinsic ethnic factor** and the latter as **an extrinsic ethnic factor.** Ethnic factors are organized in the ICH E5 guideline "Appendix A: Classification of intrinsic and extrinsic ethnic factors" (Figure 4)[10]. These concepts have been used for a country's regulatory submission to interpret clinical data packages from global simultaneous development or from bridging studies which include results from clinical trials conducted separately in foreign countries. The intrinsic and extrinsic ethnic factors listed in Figure 4 are general examples; whether these factors influence the therapeutic effect of a particular drug is a separate issue.

[9] European Medicines Agency (EMA), Committee for Medicinal Products for Human Use (CHMP), Guideline on the investigation of subgroups in confirmatory clinical trials., 31 January 2019; Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-investigation-subgroups-confirmatory-clinical-trials_en.pdf (last accessed 20 May 2022)

[10] (Ministry of Health and Welfare, Director-General of the Pharmaceutical Safety Bureau Notification, Ethnic Factors to Consider When Accepting Foreign Clinical Data, Pharmaceutical Affairs Council No. 672, August 11, 1998 (https://www.pmda.go.jp/files/000156571.pdf; last access confirmed May 20, 2022).

APPENDIX A

Classification of intrinsic and extrinsic ethnic factors

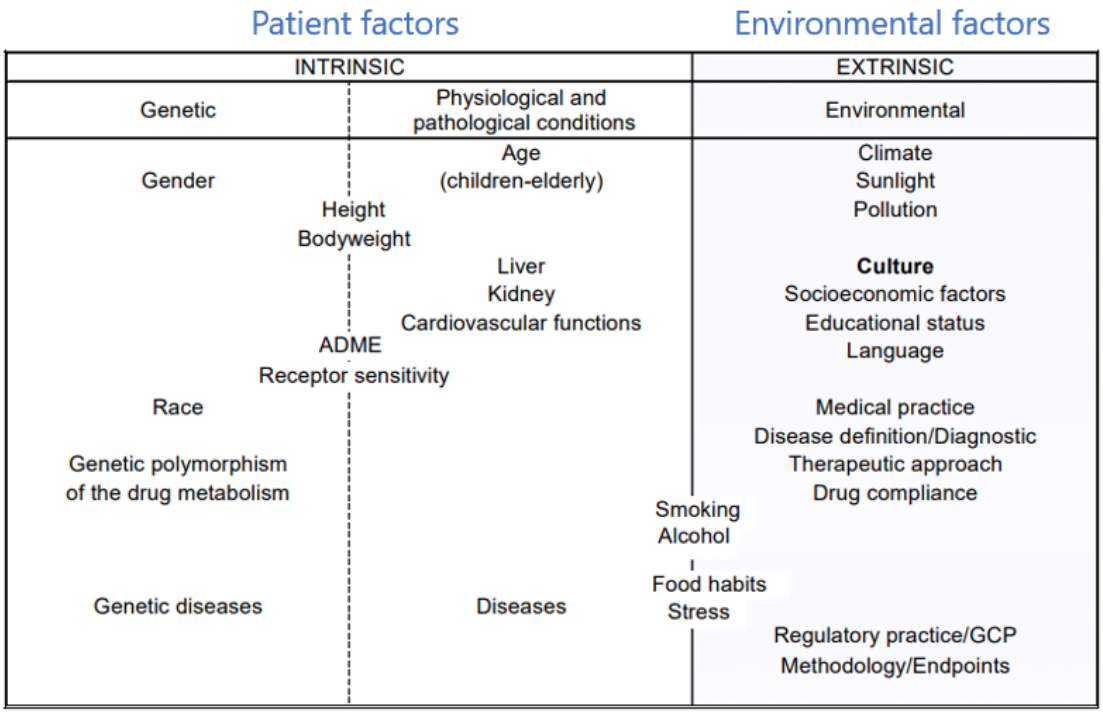


Figure 4 ICH E5 Guidelines Addendum A: Classification of intrinsic and extrinsic ethnic factors (Headings of Patient Factors and Environmental Factors are added for explanation)

In this report, we introduce the term **effect modifier,** which adds the perspective of whether a particular drug affects the treatment effect to the intrinsic and extrinsic ethnic factors. By rearranging the various factors listed in Figure 4 on the new axis of "strength of evidence supporting an effect on treatment effect (effect modifier described below)," factors are classified into the three categories; factors that have been established as effect modifiers based on sufficient evidence, factors that have not yet been established but are still considered as candidates of effect modifiers, and factors for which no evidence of effect modification has been obtained (Figure 5). Factors for which no evidence supporting effect modification has been obtained may include potential effect modifiers that are unknown in the course of the development program or at the time of submission for approval. To address this issue, it is important to evaluate from a regional or racial perspective that is larger than a country level. The discussion in this report is mainly in the context of global development. However, the discussion in this report is not limited to global development, but can be applied to various cases including:

- Evaluation of drugs developed exclusively in countries with limited ethnic diversity (e.g., Japan) and seeking approval in those countries
- Development of drugs for rare diseases for which it is difficult to conduct confirmatory studies
- Development of drugs for special patient populations such as children and the elderly
- Model-informed drug development (MIDD)

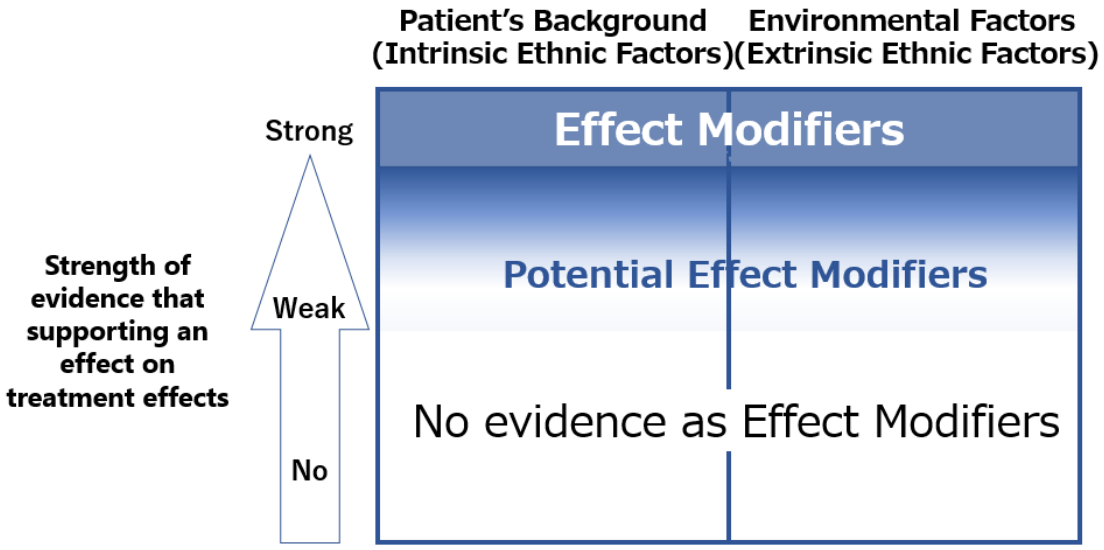


Figure 5: Relationship between patient and environmental factors and effect modifiers

In general, a biological phenomenon in which the treatment effect (Note) differs depending on the situation is called **effect modification**, and the factors that cause effect modification are called **effect modifiers**[11] (Figure 6).

(Note: A more generalized definition that includes epidemiology may be **effects of exposure**.

[11] Boston University School of Public Health, Effect Modification (https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_multivariable/BS704_ Multivariable4.html; last access confirmed May 22, 2022)

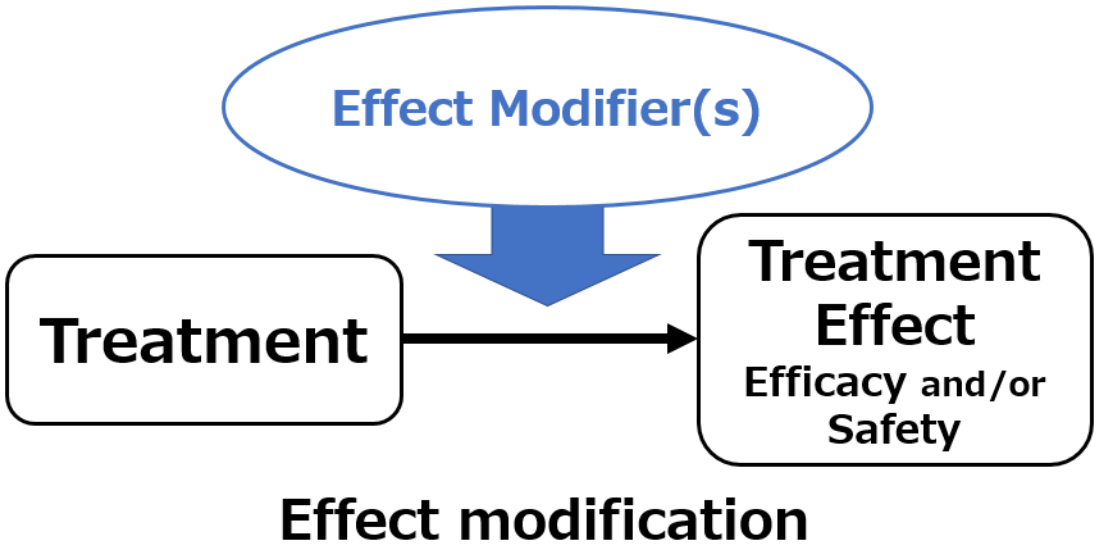


Figure 6 Effect modifiers and effect modifiers

Column: Effect Modifiers and Related Terms

Effect modifiers are defined here based on information available prior to the start of treatment, although new effect modifiers may be identified as more information on the treatment effect is accumulated. The main reason is that the information must be available at the time the physician decides to treat the patient, otherwise it cannot be used to predict the treatment effect for the patient. Prognostic factors of the disease may be effect modifiers, while patient or environmental factors that are not known as prognostic factors may be effect modifiers. In the safety assessment, factors that increase the risk of a given adverse drug reaction are called risk factors, which are also effect modifiers according to the definition above. In the case of adverse drug reactions that are related to the biological pathway for efficacy to occur, efficacy and safety may be affected by the same effect modifier, and for adverse drug reactions that occur by a different mechanism than efficacy, the effect modifier may be different from those for efficacy. For example, HLA antigens associated with increased risk of Stevens-Johnson syndrome in patients exposed to phenytoin or carbamazepine do not contribute to effect the modification for efficacy[7].

A simple and typical method for identifying the presence of effect modifiers is to divide the population into subgroups according to some patient or environmental factor and examine differences in treatment effects (Figure 7). If the treatment effect differs among subgroups,

this suggests that the factor may be an effect modifier. The procedure for finding candidate effect modifiers is discussed in detail in the next section.

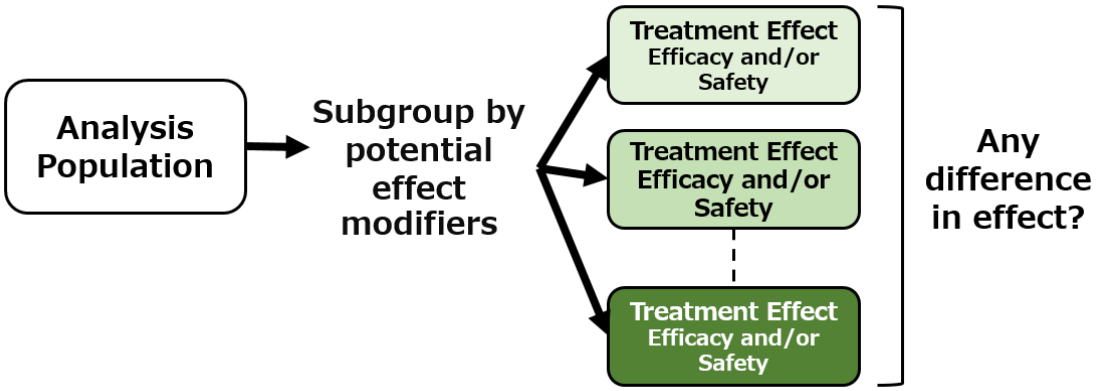


Figure 7: Subgroup analysis for effect modifiers

Finding Candidate Effect Modifiers

There may be a very large number of effect modifiers for each drug and even for each endpoint. However, it is not realistic or practical to consider every effect modifier, including those that have only a trivial effect. Our goal in drug evaluation is to focus on effect modifiers that can explain well the differences in among people and that will be available when the drug is approved for marketing and becomes widely used in the medical field. Hints leading to potential candidates for such effect modifiers may be obtained throughout the drug discovery research phase and clinical development. For example, various molecules and measurable biomarkers involved in the biological response processes associated with the pathological conditions targeted by the drug, as well as known genetic polymorphisms among them in humans, may harbor candidate effect modifiers that explain the differences in response among individuals in humans. Efforts to identify candidate effect modifiers will continue in clinical development. However, in early development (especially before Proof of Concept studies), major investment decisions for the later stages of development have not been made, the evaluation that forms the basis for those decisions is a top priority, and budgets are often not allocated to allow for sufficient "nice-to-have" studies. In other words, the focus is on whether the candidate drug is likely to have the expected efficacy and safety profile in humans, and whether there are differences in response in specific countries or regions is a secondary issue to be considered. Therefore, in early development the search for candidate effect modifiers will continue to focus on patient factors rather than environmental factors.

The process from drug administration to the expression of the treatment effect (efficacy and safety) consists of the drug concentration in the body after drug administration, the pharmacodynamic effect caused by the drug, and the expression of treatment effects, each of which is interrelated.

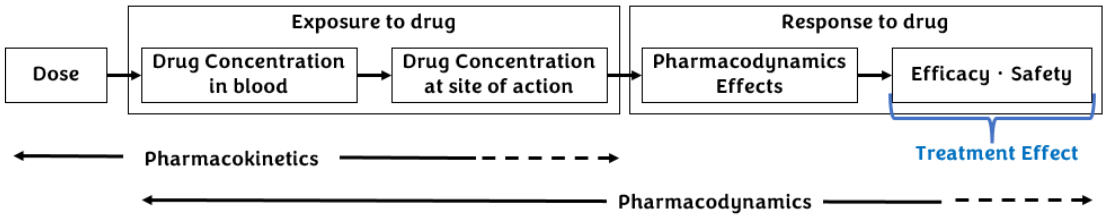


Fig. 8 Relationship between drug dose, exposure, and response ("Guideline for Exposure-Response Analysis of Drugs" Fig. 1: Notification No. 0608-4 dated June 8, 2020 by the Director, Drug Evaluation and Control Division, Pharmaceuticals and Vital Health Bureau, Ministry of Health, Labour and Welfare)

Interindividual differences in exposure are due to interindividual differences in absorption, distribution, metabolism, and excretion. Patient factors generally include age, gender, body size (weight, BMI), liver and kidney function, and genetic polymorphisms of metabolic enzymes, while drug interactions and dietary influences are often considered as environmental factors.

For example, much knowledge has been accumulated on the effects of gene polymorphisms on the activity of cytochrome P450 (CYP) genes involved in the metabolism of many low-molecular-weight drugs and on the frequency of gene polymorphisms in different races. This knowledge has been accumulated from a variety of sources, including in vitro studies such as metabolism experiments, results of clinical trials of many drugs, and genetic analysis (see column). For antibody drugs, for example, some examples have been reported in which population analysis models of various drugs were examined and general factors affecting pharmacokinetics were discussed.[12]

Column: Genetic Polymorphisms of Metabolic Enzymes and Transporters

[12] Bensalem, A., Ternant, D. Pharmacokinetic Variability of Therapeutic Antibodies in Humans: A Comprehensive Review of Population Pharmacokinetic Modeling Publications. *Clinical Pharmacokinetics*. (https://doi.org/10.1007/s40262-020-00874-2; last access confirmed May 20, 2022)

For response, factors that may be prognostic or predictive of treatment, such as disease status, are often explored as patient factors, and diagnostic criteria, treatment, and concomitant medications are often considered as environmental factors.

Throughout clinical development, efforts to identify candidate effect modifiers and their impact on response (i.e., the extent to which factors can explain a variation in response) will continue in individual clinical trials. A typical approach to summarizing the results of individual trials is described in section "A) Adding Insight to Existing Information" of **Application of Effect Modifiers**. As clinical development progresses, the search for effect modifiers will expand beyond patient factors to include environmental factors such as diagnostic criteria, treatments, concomitant medications, and other medical environment factors.

To understand the degree of influence in this process, it is important to compare populations characterized by the presence or absence of factors. Even at the stage where factors cannot be identified, differences in response to treatment can be seen by comparing large groups that include various factors such as race and country (e.g., when the distribution of genetic polymorphisms in one race is different from that in another race, or when everyone in one country is affected by the same extrinsic ethnic factors). The broad cut includes factors such as race and country. It is important to maintain the perspective that there may be unknown effect modifiers lurking in the broad cut. These studies are ultimately examining factors that are predictive of an individual patient's response.


Column: Exploring Effect Modifiers -Alogliptin Example

Further Exploration of Effect Modifiers

The identification of effect modifiers is not something that can be proven using a confirmation framework, as described in **Fundamentals for Improving Evidence**. Rather, it is a process of building confidence by accumulating evidence from various findings in the development program supporting that a factor is indeed an effect modifier. As this learning process continues, information on multiple candidate effect modifiers accumulates. The effect modifiers or potential effect modifiers may be narrowed down based on the magnitude of the effect of each factor on the treatment effect, correlations among factors, and ease of use in clinical practice.

Depending on how many effect modifiers and their candidates are captured collectively, including potential factors, we can think of the hierarchy as shown roughly in Figure 9. For illustrative purposes, this figure also includes ethnicity, which is the aggregation of many ethnic factors, and geographic region/regulatory region, which is a much larger perspective.

Moving up the pyramid in Figure 9 is one aspect of the learning process for effect modifiers, and the model is updated along with the learning process. This learning process does not always have to start at the lowest level and climb up one by one, but can start at any level. The cases where, in the drug discovery research stage, candidates for factors included in the top hierarchy (e.g., genetic polymorphisms of molecules important in the mechanism of action) are identified and confirmed in clinical trials, would increase. Effect modifiers (or candidate effect modifiers) may be identified based on known information or in early-stage development trials.

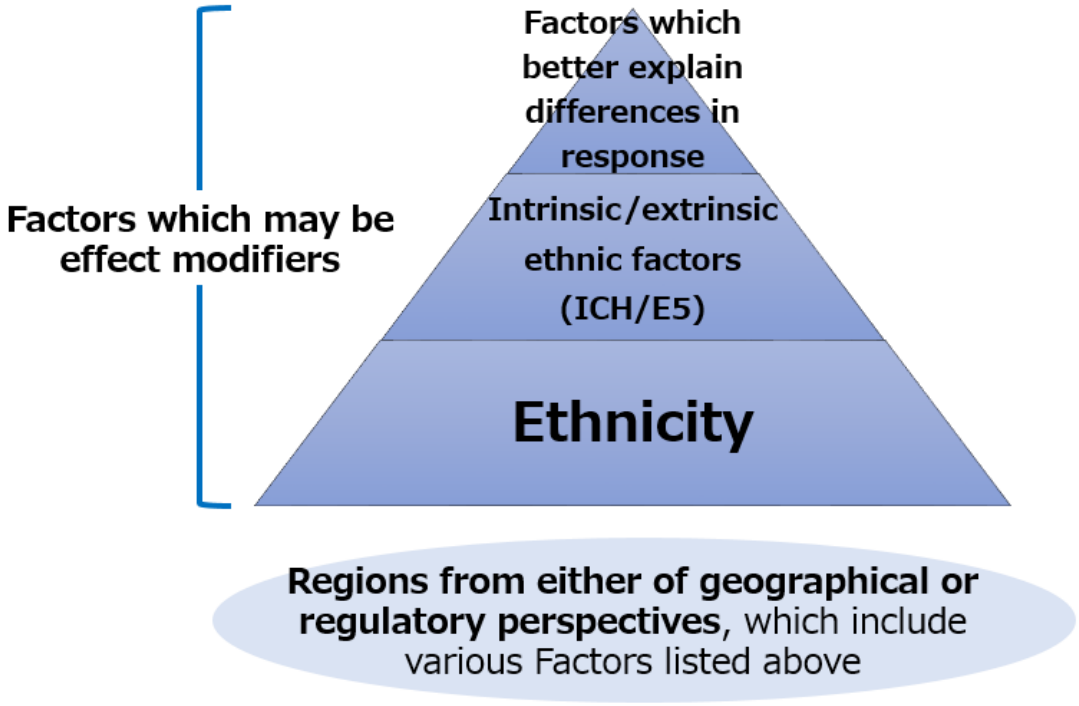


Figure 9 Various effect modifiers

The discovery of a candidate effect modifier does not necessarily mean that the same effect modifier will continue to be tracked in the subsequent learning process of drug evaluation. There may be multiple, more detailed effect modifiers behind the identified effect modifiers, and an effect modifier may be recognized as an observable phenotype. For example, body weight is a patient factor (intrinsic ethnic factor) and is often considered an effect modifier. If body weight is easy to measure and useful in interpreting and estimating treatment effects, it may have great clinical utility, and weight may be used as an effect modifier. However, genetic background, comorbidities, lifestyle habits, or environmental factors that are closely related to body weight may provide sharper explanations of their effects on treatment effects. In such cases, these effect modifiers would replace body weight with newly identified factors.

As described in **Finding Candidate Effect Modifiers**, starting with an exploration for intrinsic and extrinsic ethnic factors is a typical approach. Once we have a story in the "intrinsic and extrinsic ethnic factors" hierarchy that can explain differences in treatment effects, we may proceed to explore "further specified factors," a higher hierarchy. Knowing the distribution of the identified effect modifiers in a population of interest, we may be able to more rationally explain differences in treatment effects in the lower hierarchies of "ethnicity" and "geographic region/regulatory region."

On the other hand, in some cases, whether early or late in development, not even candidate effect modifiers have been identified. One strategy is to intentionally include a wider range of ethnicities and geographic regions in the trial so that the effect modifiers can be examined post hoc from many angles as possible. Whether to pursue such a strategy, and if doing so, how broadly, is an important and difficult decision for both the company developing the drug and the regulatory agency, but the determining factor is "how can the treatment effect be explained in the patient population in the country or region where approval is sought?


Column: Exploring Effect Modifiers -Gefitinib Example

Application Scenarios for Effect Modifiers

What are the benefits of effect modifiers when they are identified? There are two major opportunities for effect modifiers to be useful.

[Adding insight to existing information]

Explain the results obtained from a study or information that is already available using effect modifiers to help understand the treatment effect of a drug.

[Helping to predict the future]

Provide useful information for future research planning based on existing knowledge of effect modification and for selecting treatment strategies by predicting treatment effects for individual patients in future clinical settings.

Let's discuss these opportunities in more detail.

A) [Adding insight to existing information]

Looking back at previous practices related to effect modifiers, we have shown the presence of effect modifiers through subgroup analysis as shown in Figure 7 by various factors (demographic data, prognostic factors of the disease). For example, suppose that the severity of the disease being treated is a candidate effect modifier. If severity of disease is on the horizontal axis and a treatment effect is on the vertical axis, the relationship between the severity of disease and the treatment effect of individual subjects can be represented as a scatter plot on this plane (Figure 10). If a relationship is found between severity and treatment effect, e.g., showing patients with more severe disease are more likely to have a higher treatment effect, this would suggest that severity is an effect modifier. Figure (2a) of the E17 Guideline[13] conceptually shows the oval-shaped region that contains the majority of the

[13] (Notification of the Director of the Pharmaceuticals Evaluation and Management Division, "General Principles for the Planning and Design of Global Clinical Trials," Pharmaceutical and Pharmaceutical Affairs Bulletin No. 0612-1, June 12, 2018 (https://www.pmda.go.jp/files/000224557.pdf; last access confirmed May 20, 2022).

plots for each subject. This figure illustrates how the distribution of responsiveness across regions can look different if the distribution of severity, an effect modifier, differs across regions, even though there is a tendency for patients with greater disease severity to have a higher treatment effect across trials and the trend is the same across regions.

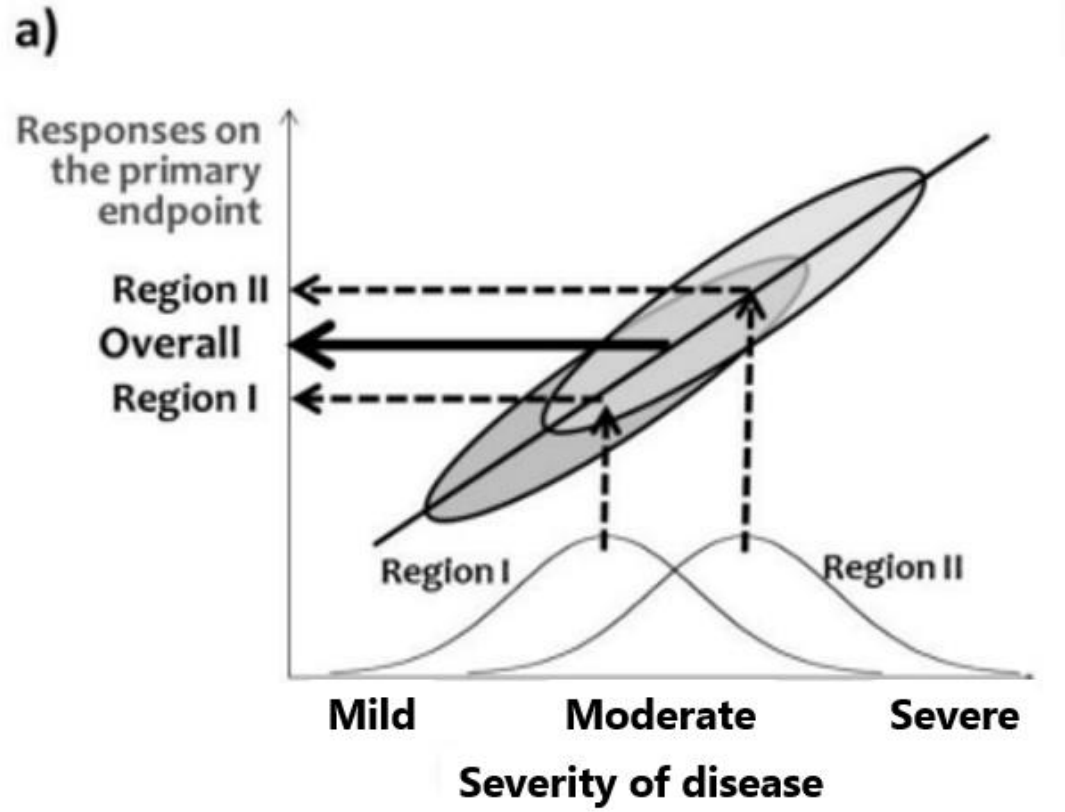


Figure 10 Relationship between the distribution of effect modifiers and treatment effect (adapted from Figure 2a in the text of the ICH E17 Guideline).

An explanation based on the approach shown in Figure 10 may be easier to understand when there is only one effect modifier. When there are two or more effect modifiers (e.g., when weight is an effect modifier in addition to severity in the example above), there are several possible approaches to explain effect modification. Some of these are listed below.

A1: One factor at a time approach

This approach repeats the explanation based on the subgroup analysis as shown in Figure 7 for each candidate effect modifier. This approach is the simplest and allows one to examine the relationship between each effect modifier and the treatment effect, but it cannot explain the combined effect of multiple effect modifiers. If the effect of each effect modifier on the treatment effect is independent of the other effect modifiers, the combined effect of multiple

effect modifiers will be the sum of the effects of the individual effect modifiers. The conclusions obtained by clarifying the relationship between each effect modifier and the treatment effect are the same as those obtained by the following approaches (A2, A3).

A2: Approach to subgroup analysis with a combination of effect modifiers

This approach attempts to account for the effects of each effect modifier alone as well as the combined effects of multiple effect modifiers. For example, suppose the case where the severity and weight of the target disease are effect modifiers. If the severity of the disease is divided into three levels (mild, moderate, and severe) and the weight into three levels (low, normal, and high), there will be nine subgroups based on the combination of the two effect modifiers. By looking at how treatment effects differ among these nine subgroups, the approach attempts to explain the combined effects of severity and weight. One concern is that the larger the number of subgroups do not provide sufficient information and may be more difficult to interpret.

A3: Approach using statistical models

Since the statistical model can incorporate multiple effect modifiers simultaneously, it is also possible to examine the effects of each effect modifier and their combined effects, i.e., interactions among effect modifiers. In subgroup analyses such as A1 and A2, it is necessary to consider each effect modifier in ordinal categories with a threshold set somewhere for each, but in statistical models, effect modifiers can be treated either as ordinal categories or as continuous quantities, and interactions can be examined in either case.

B) [Helping to predict the future]

Incorporating findings on effect modifiers into new study designs is an example of this application. When there is evidence from past studies to support that some factor is an effect modifier, it is conventionally used as a stratification factor during subject allocation or included in statistical models as part of the explanatory variables (covariates) in the main analysis. (Note)

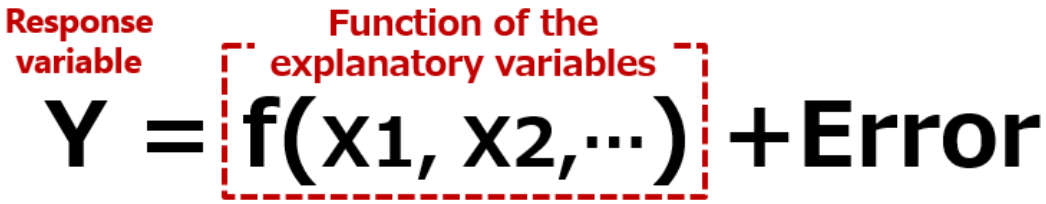
(Note: Such application scenarios include cases where the evidence for effect modification is insufficient, and the factor is just a candidate for an effect modifier. When there is fairly clear evidence for effect modification, enrichment may be performed to narrow down the subject population in terms of effect modifiers. A situation in which we use a candidate

effect modifier as a stratification factor or covariate may be a situation in which we expect predictability of effect but have not yet obtained enough evidence of effect modification to decide to enrich the population. There is no general, clear threshold for the weight of evidence for effect modification, nor for the decision to enrich or use as a stratification factor or covariate. How to handle effect modifiers and their candidates will depend on the sponsor's development and submission strategies and will also involve the judgment of regulatory agencies on issues that require their agreement.

In a MRCT, this could be applied to a prespecified pooling strategy recommended in the E17 Guideline or to an exploratory analysis plan using a pooling strategy. The pooling strategy in MRCTs can help predict treatment effects for a patient population characterized by effect modifiers (or their distribution). This can be taken further to predict treatment effects for individual patients characterized by effect modifiers. What makes this possible is the elaboration of the model. By incorporating effect modifiers directly related to individual patient characteristics, such as demographic data, degree of disease progression, and detailed patient background such as genetic information, rather than broadly defined factors such as geographic region, country, and race, the model can quantify the treatment effect on individual patients in future medical practice.

Column: What is a Region and What is a Consistent Evaluation of Treatment Effects across Regions?

In general, a statistical model expresses the **response variable** (Y), such as the treatment effect or risk of event occurrence, as a sum of a function of the **explanatory variables** (x_1 , x_2 , ...), such as treatment, patient factors, and environmental factors, and residual **error**.



The function of the explanatory variable gives the average value (point estimate), while the error expresses the variation around the average value. Error is a relative concept, meaning that variables other than those included in the function of the explanatory variables are not explicitly considered as contributing factors to the response, and the effects of the various

factors are all considered together as errors and are assumed to follow a specific probability distribution. In addition, the function of explanatory variables may include interactions among factors.

B1: If **there are no known effect modifiers,** the only explanatory variable is the treatment (including the dose of treatment), and the statistical model is

$$Y = f(\text{Treatment}) + \text{Error}$$

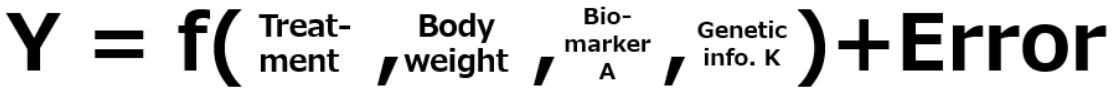
The value of the f(Treatment) part is determined by whether the patient received the treatment or not, regardless of any factors rather than treatment. Though the response differs from country to country, region to region, or person to person, the explanation is that those are errors.

***B2*: If no effect modifiers can be identified, but the response is expected to differ across regions (a set of one or more countries)**, then treatment and region are the explanatory variables and the statistical model is

$$Y = f(\text{Treatment}, \text{Region}) + \text{Error}$$

The f(Treatment, Region) part determines the value depending on whether the person received the treatment and in which region the person lives. The explanation is that within each region, the response may differ from person to person, but this is an error. It is also possible to add a treatment-region interaction term to the function of the explanatory variables, and to consider whether the interaction is qualitative (the treatment effect is opposite in each region) or quantitative (the treatment effect is the same in each region, but the magnitude is different). This is an important consideration for subsequent development plans and regulatory filings.

***B3*: If effect modifiers are identified and they are modeled**, then those effect modifiers in addition to the treatment become explanatory variables, and a statistical model can be used. For example:



The part f(Treatment, Weight, Biomarker A, Genetic information K) is the average value for an individual who received with a combination of three effect modifiers: the specific body weight, biomarker A (which may be the measurement itself or a category with some classification), and genetic information K (which may be the presence of a specific mutation or a category such as a genetic polymorphism). The error is the remaining variation that cannot be explained by f (Treatment, Weight, Biomarker A, Genetic information K). Effect modifiers can be patient or environmental factors. The more effect modifiers that can explain the differences in response are modeled, the smaller the error becomes, in the order B1, B2, and B3 in the example above. If such modeling can be done in a development program, the probability of trial success can be estimated by simulation when considering a new trial design. For example, if the severity of the target disease is an effect modifier, multiple scenarios can be set up for the percentage of subjects with mild/moderate/severe composition, and the probability of success for each scenario can be calculated to examine the trial design. The more the combination of explanatory variables in the model can represent the background of each individual patient, the more the functional part of the explanatory variables (f [treatment, weight, biomarker A, genetic information K in the example above]) can give a predictive value of the treatment effect for each individual patient, together with errors representing uncertainties that cannot be modeled. The presentation of treatment effects can provide important information for treatment selection in the medical practices.

Column: Using Effect Modifiers to Predict the Future - A Personalized Medicine Approach

There may be situations where the approach described in section **B) [Helping to predict the future] will be** useful in **A) [Adding insight to existing information]**, i.e., in the analysis and interpretation of trial results. The approaches described in A1, A2, and A3 above may be attempted using candidate effect modifiers suggested by the trial, and the explanation of the treatment effect may be strengthened by the candidate effect modifiers. In a MRCT, candidate effect modifiers could be used to construct post hoc pooled regions or pooled

subpopulations. If differences among these populations are observed, it would suggest that the factor is an effect modifier.

Explanation by effect modifiers also helps regulatory bodies to review all the information obtained in the development program and to make a decision on whether to approve or disapprove the drug for use in the patient population in the country. Regulatory decisions are challenging in that they involve predicting treatment effects in future patient populations for which data are not yet available, but predictability may be enhanced if effect modifiers are identified and the distribution of effect modifiers in the patient population in the country is known. It would also help to predict what proportion of the patient population in the country (or any patient population characterized by effect modifiers) is likely to benefit most and least likely to benefit from the treatment.

So far, there are few examples of models that have been constructed to predict the treatment effects of drugs as described here. There have been cases where the effects of drugs have been examined as part of models to predict disease prognosis, such as in osteoporosis[14] and breast cancer,[15] and models have been devised to predict the risk of various events in the Framingham study.[16, 17] These examples provide us with important clues. We hope that various attempts will be made to construct predictive models from the perspective of drug efficacy and effect modification, and that the learning process leading to the construction of models useful in clinical practice (including updates of predictive models based on new information) will be widely shared.

[14] N.C. Harvey, FRAX and the effect of teriparatide on vertebral and non-vertebral fracture, Osteoporos Int 26(11):2677-84 (2015).

[15] Candido dos Reis et al., An updated PREDICT breast cancer prognostication and treatment benefit prediction model with independent validation, Breast Cancer Research 19-58 (2017)

[16] Hirai, H. et al., New risk prediction model of coronary heart disease in participants with and without diabetes: Assessments of the Framingham risk and Suita scores in a 3-year longitudinal database in a Japanese population, Scientific Reports volume 9, Article number: 2813 (2019)

[17] Lloyd-Jones, D.M., Framingham risk score and prediction of lifetime risk for coronary heart disease, Am J Cardiol 94(1):20-24(2004).

What Can Be Done?

Current Issues and Historical Background

Evaluating the efficacy and safety of a drug in a country's patient population based on data actually collected in that country has been considered convincing, and is a view shared by many people in several countries, including Japan, even today. This is due to the historical background of the global simultaneous drug development.


Column: History of Drug Development - From Local Development to Global Simultaneous Development

The global simultaneous drug development gradually progressed from the mid-2000s, and data collected in foreign countries became an increasing part of the clinical data package subject to review in Japan. In line with this trend, there were an increasing number of cases in which data collected in Japan were insufficient not only for the confirmation of treatment efficacy in the Japanese population alone, but also for the secondary evaluation of efficacy and safety evaluation in the Japanese population alone. The same problem exists for drugs for rare diseases and special populations, regardless of whether they are developed through global or domestic drug development. However, even under such circumstances, an approach that places results from the Japanese population vs. results from the overall or non-Japanese population at the core of data presentation and data evaluation (referred to here as the "E5-like approach") has been used to evaluate the efficacy and safety of a drug in the Japanese population. The E5-like approach has been applied to many situations in drug development, including the following:

- MRCTs Including Japanese Population
- Japanese Phase I study prior to participation in MRCT
- Long-term safety study in a Japanese population in a global simultaneous development program
- Additional studies in Japanese populations in situations where efficacy and safety profiles are somewhat well understood from foreign data

Companies that consider the E5-like approach to be de facto standard (de facto norm) may ask, "We understand it is a good practice to conduct trials like the ones above, and if possible, can we skip these trials? And if possible, can we reduce the number of Japanese cases?" However, we would like you to reconsider this once again. In various situations of drug development, the E5-like approach has itself become the starting point for discussions on how many Japanese cases are needed, considering that Japanese data are essential to obtain marketing approval. Discussions based on these approaches do not have a high affinity with the approach of trying to explain treatment effects such as efficacy, safety, and PK/PD from the viewpoint of effect modifiers, as discussed in this report. There are whispers in the industry that "Our company's arguments are not easily accepted by the regulatory authorities" or "Regulators are stubborn no matter what we say and apply existing notices by the book," but is this really true? It seems that the problem lies in the communication between companies, academia, regulators, healthcare professionals, and patients, each of whom are considering different outcomes from their own standpoint or the "setting of the game" itself Thinking that the person you are discussing or communicating with is a "know-nothing" is a quick way out to justify oneself. Regulators and sponsors have different positions, but are we discussing the same scientific playing field? In such a debate, could we have developed a story that could be explained to any regulator without jumping to "Japanese vs. …" comparisons? Could we have constructed an explanation that would be acceptable not only to the Japanese regulators, but to any regulator in the world? Depending on what the starting point for discussion is and how the explanation is constructed, the content of the consultation meeting with regulatory authorities in the planning phase and the story of the CTD should be different. These would be root causes for that there has been little progress in understanding and implementation of the E17 Guideline, several years after the E17 Guideline was agreed upon at the ICH in 2017. We should carefully consider the significance of the E17 Guideline being agreed upon as a separate guideline, rather than as a revision of the E5 Guideline itself or E5 Q&A No. 11, which discussed MRCTs within the framework of the E5 Guideline.

Today, official English translations of the Japanese review reports are available[18] to provide and promote the use of Japanese review information to other countries. The same applies to the package inserts, which are referenced by many countries. In the case of drugs undergoing global simultaneous development or being granted marketing authorization in Japan earlier than in other countries, Japan is now at the

[18] Specifications for English Translation of FY2021, FY2022, and FY2023 Examination Reports (https://www.pmda.go.jp/files/000238622.pdf; last access confirmed May 20, 2022)

forefront of evaluating the efficacy and safety of drugs and is responsible for disseminating information that is useful to other countries and contributes to appropriate uses of drugs around the world.


Column: Argument Based on Data From a Small Number of Cases


Column: Who is Japanese?

Currently, however, evaluations are made based on the results of cutting out information on the Japanese population, which is part of the overall population of a MRCT, and comparing it with the overall results. Disseminating such data and evaluation results overseas would not provide useful information for regulatory authorities in other countries or people in the medical environment in other countries to predict treatment efficacy in their own patient populations. Regulatory authorities in each country should consider the benefits & risks in each country's patient population and whether it is acceptable to the society in that country before granting marketing authorization. When foreign regulatory authorities judge whether the results of the approval review in Japan can be applied to their own countries, they would consider whether environmental and patient factors in Japan and their own countries are similar. Again, more detailed information on effect modifiers and their impact on treatment effect would allow foreign regulatory authorities to make decisions based on scientific considerations, such as whether to grant approval. The approach of using effect modifiers to estimate the treatment effect in the patient population of interest, based on data from around the world and information external to the development program, could be shared by regulatory authorities in any country. It would provide an easy-to-understand explanation of the benefits and risks of the drug in question for any country. From the perspective of explaining the benefits and risks to the public not only in Japan but also in other countries where approval is sought, it would be more persuasive to discuss the benefits and risks using effect modifiers such as patient and environmental factors, based on data and information obtained worldwide. If the information becomes more detailed to the point of knowing what type of a patient can be expected to have a higher level of efficacy and an increased risk of clinically significant adverse drug reactions, it will be useful information for both the physician addressing the patient and the individual patient to consider the benefit-risk relationship on

an individual patient basis. Now is the time to rethink the conventional E5 approach and think carefully about what kind of thinking should be used in drug evaluation.

Column: What We Have Seen in the Coronavirus Pandemic

What can we do to prepare ourselves for such a future?

What is needed is the following approach to drug evaluation.

STEP 1:	**To identify effect modifiers in development programs based on data collected worldwide, or to maintain such a perspective.**
	The learning process for effect modifiers begins in the nonclinical research phase. If human genetic polymorphisms are known for molecules that appear in the story of a drug's mechanism of action, they may be candidates for effect modifiers that explain regional and ethnic differences. There may also be candidate effect modifiers among known prognostic factors for the disease being treated and environmental factors that vary from country to country and region to region. We will attempt to construct a statistical model to predict treatment efficacy, and safety if possible, with these candidates in mind. The statistical model for predicting treatment effect will be updated by taking into account data collected during clinical development and new external information such as literature.
STEP 2:	**Characterize the patient population of interest (country/region, demographic group) with effect modifiers.**
	In other words, it is to know the distribution of effect modifiers in the patient population of interest. The source of this information may come not only from data collected in the development program, but also from the literature and national or global epidemiological data.
STEP 3:	**Describe the treatment effect in a given country/region or in a given patient population based on a predictive model using effect modifiers with estimable logic.**
	This explanation will allow us to move away from an interpretation that relies on data from a small number of cases actually collected in a clinical

	trial and allow us to base our discussion on an estimate of the treatment effect.

The same basic approach can be taken for safety evaluation. Unlike efficacy evaluation, in which the endpoints which should be used are already determined, the safety evaluation always requires the following considerations in addition to STEPs 1, 2, and 3 above.

Safety Evaluation	**To the extent possible, the overall picture of what kind of adverse events may occur and the causal relationship to the drug product should be examined. If necessary, measures will be taken to minimize the risk in the development program.**
	First, adverse events are comprehensively collected to determine what happens to people who use the drug. For each adverse event reported in an individual case or multiple cases, a causal relationship to the drug is examined based on the results of one or more trials. If the evidence supporting a causal relationship with the drug is strong or weak, and it is deemed necessary to reduce the risk to subjects, measures will be taken, such as reflecting this in the inclusion criteria for the clinical trial or carefully collecting and reviewing relevant information.

For pharmaceutical companies developing with an eye on the global market, especially mega pharma with large pipelines, the main interest early in a global development program (up to the Proof of Concept study) is not patient or environmental factors in a particular country or region, but rather "What happens when a drug is administered to humans?" In order to efficiently answer this question and to avoid large investments in early development, sponsors have adopted a strategy that emphasizes the accuracy of pharmacokinetic and pharmacodynamic data, and in Phase I and early Phase II trials, sponsors have focused on a narrow range of countries and sites in which to conduct trials. In Phase I and early Phase II trials, sponsors tend to narrow down the countries in which the trials are conducted and narrow down the number of sites, while also examining safety and tolerability. In later stages of development programs, the number of participating countries and sites is increased and the focus shifts to evaluating efficacy and safety in a population with a variety of patient and environmental factors. The basic concept of safety evaluation in such global development programs is as follows:

The safety information is viewed from a global perspective, and when an imbalance is found between populations characterized by some factor (an effect modifier for an individual event, sometimes called a risk factor in the context of the safety assessment), such as frequency of occurrence or distribution of severity, research is performed for the cause of the imbalance.

When there is a clinically significant adverse event and the causal relationship to the drug is uncertain or the factors that increase the risk (patient and environmental factors) are not known, it should be considered a risk to humans all over the world, and that the same level of risk exists in the country where the event was reported as does in countries where it was not reported. If such a view is accepted, then all countries participating in a global simultaneous development program stand together at the forefront of the drug's evaluation and share in the unknown risk, and there is little rationale for an approach that requires particular countries to be particularly cautious or to consider first what might happen to their own patient populations.

The ideas described so far may be applicable to the development of drugs for special patient populations such as children and drugs for rare diseases. In the development of drugs for special populations and rare diseases, it is not uncommon for the size of the patient population to be treated to be small and, due to ethical considerations, for the number of subjects to be included in clinical trials to be small. In such situations, it would be more scientific to build a predictive model that does not strongly rely on data from a small number of subjects actually collected, but also takes into account information external to the development program, and to conduct the drug evaluation while visualizing the uncertainty of the prediction.

For example, let us assume that a drug has been developed globally for adults and that a somewhat reliable predictive model for treatment effects in adults has been developed. When the same drug is developed for pediatric patients, the pediatric predictive model may need to be updated to reflect the adult model. The update will consider whether the update can reflect universal adult and pediatric differences, or whether it should be treated separately depending on country- and region-specific environmental and patient factors. The ICH E11A Guideline (Extrapolation in Pediatric Drug Development)[19] builds on this thinking and discusses similarity of drugs, similarity of response to treatment, and usage of a variety of

[19] Draft ICH E11A Pediatric Extrapolation (https://www.ich.org/page/efficacy-guidelines; last access confirmed June 9, 2022)

existing data. It is the understanding of effect modifiers and the predictive models that make this visible that allow us to estimate benefit-risk in a country's pediatric population without relying on data obtained from a small number of children in their own country.

If we assume the typical data collected in conventional drug development, the approaches discussed in this report may not seem feasible. However, it is too pessimistic to assume that this situation will continue for a long time to come. Compared to the timeframe that typically takes more than 10 years from the discovery of a new drug to its development and then to its marketing approval, analytical and measurement technologies such as imaging, Omics, and mobile devices advance very quickly and are indeed advancing day-by-day. The key to the approach discussed in this report is the understanding of effect modifiers. If we can further deepen our understanding of patient and environmental factors and identify factors at the molecular and genetic levels, the search for effect modifiers will go much deeper than in the past. We have been discussing this future with this in mind, and the European Organization of Research and Treatment of Cancer (EORTC), a European cancer-related research organization, has made an interesting proposal that is highly compatible with this report and suggests an exciting future. Figure 11 is a conceptual diagram modified from a figure published in the EORTC paper.[20]

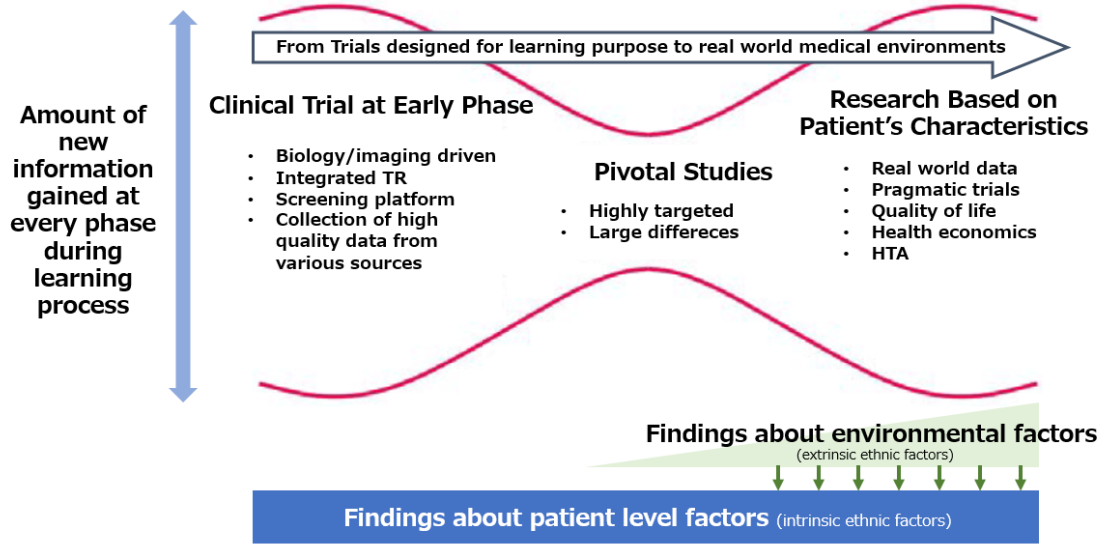


Figure 11 Diabolo-shaped approach (aerial sesame-shaped approach)

[20] Lacombe, D. et al. Precision Medicine: from "Omics" to Economics towards Data-Driven Healthcare - Time for European Transformation, *Biomed Hub* **2(suppl 1)**, 212-221 (2017). (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6945945/pdf/bmh-0002-0212.pdf; last access confirmed May 20, 2022)

In Figure 11, the horizontal axis represents time and the vertical axis represents the amount of **new** information obtained at each stage. Conventional drug development has taken the approach of starting with a small number of subjects to be evaluated, increasing the number of subjects and expanding to subjects with various backgrounds, and then confirming efficacy and drawing safety profiles in confirmatory trials ("Pivotal Studies" in Figure 11) before submitting a drug for approval, with an element of half-gambling. However, just like the use of longitudinal data obtained from imaging, Omics, and mobile devices, the amount of information obtained from each subject will increase by an order of magnitude, and there will be more examples where effect modifiers can be explored at an early stage of development without increasing the number of subjects as compared to today. As our understanding of effect modifiers deepens and models mature by the time we plan pivotal trials, we will be able to consider the approach of conducting pivotal trials that are relatively small by narrowing the target population and have a high probability of success, with the goal of detecting large effect sizes. Such an approach is also expected to contribute to individualized medicine. In the early stages of development, the focus will be on examining patient factors (intrinsic ethnic factors), and from pivotal trials to postmarketing, the focus will shift to examining environmental factors (extrinsic ethnic factors). The exploration of country- and region-specific environmental factors will be insufficient for many countries and regions, even if they are included in pivotal studies. The place for these studies will naturally be in postmarketing studies after approval, which will continue as studies based on the nature of the patient population in each country and region, using real-world data and taking into account the healthcare system and healthcare environment. The proposal by EORTC depicts such a future vision. Such an approach will be realized in development projects that have the opportunity to explore effect modifiers in detail from drug discovery to the early stages of development, and such development projects will gradually increase. And even if the Diabolo-shaped approach (aerial sesame-type approach) as shown in Figure 11 becomes the mainstream, the approach based on the ICH E17 Guideline and the ideas discussed in this report should not become obsolete.

Conclusion

The ultimate goal of those of us involved in drug development is to bring better medicines to patients faster. This goal, which we all will accept, should not be passed off as a slogan that is easy on the ears. What does "better medicine" mean, what does "faster" mean, what does "patient" mean - can you really see the faces of patients? What does "deliver to patients" mean? Each word of the goal has a deep meaning. There are countless paths that lead to the goal. It is our job to pave the way toward this goal, to step on what may appear to be a roadless path, and to inform patients and healthcare professionals that "this road is safe and secure," based on scientific data and considerations. This may sound a bit grandiose, but from this perspective, various questions arise, such as whether the current state of drug evaluation is the way it should be, whether it is really okay to continue with the conventional manners that have been believed to be correct, and whether we are being sidetracked by trivial issues. It is not easy to answer these questions clearly, but this report is an attempt to answer these difficult questions.

The focus of this report is on scientific thinking. Scientific thinking is what underpins our learning process and the explanations we provide to regulators, medical professionals, all those involved in health care, and the patients who are our ultimate customers. In the development of individual drugs, there are ethical considerations, business goals and constraints, and conventions that attempt to apply regulations in a prescriptive manner (in particular, conducting studies in Japanese populations and interpreting data obtained from Japanese populations), and scientific thinking alone may not be the driving force behind drug development. Even so, the challenge of drug development is to proceed while maintaining good balance with the other conditions, with scientific thinking at the core, and this is the way drug development should be, as applied science or practical science. At this point, we should not get caught up in conventions and should not use them as a starting point for discussions on individual drug development. If each convention is the result of careful consideration, including the history of why it was needed and what it was used to prevent or protect, then the likelihood of positive discussions with regulatory authorities will increase.

As mentioned at the end of the previous section, drug evaluation should be viewed from a broader perspective, rather than using the data from a small number of subjects without information obtained from imaging, Omics, or mobile devices, where no update of existing

knowledge is expected, as a key information in drug development. When you are in the early stages of development, when planning a pivotal trial, when preparing an application, or when under regulatory authority's review, please consider this as a bird's-eye view of the world's human population.

Our learning process, with the goal of "Know the drug," is like putting together a jigsaw puzzle of the complete drug profile. The overall picture will vary depending on the target patient population of the drug. The pieces of the jigsaw puzzle are the efficacy and safety factors that are relevant to the efficacy and safety of the drug, as described in this report, and the type of patients that will benefit from the drug. Please consider whether the information to be obtained in the planned trial will help fill in these pieces.

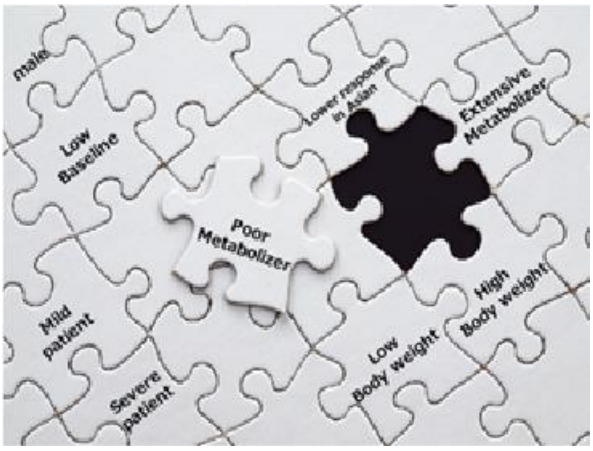


Figure 12 Components for knowing the drug

If Japan is to remain a drug-discovery powerhouse and one of the leaders in drug development in the world, we should correct the mindset of trying to explain the situation for the Japanese domestic public. Instead of the mindset of explaining to the Japanese public based on Japanese data, the mindset must be changed to explaining the same story to every regulatory authority in the world and contributing to global drug development. We hope that this report will help not only global simultaneous drug development, but also drug development for various patient populations, and that it will serve as a cornerstone for the thinking of the next generation.

FY2020 Task Force 5, Data Science Subcommittee, Drug Evaluation Committee, Japan Pharmaceutical Manufacturers Association
Osamu Komiyama (Pfizer R&D) Leader and Vice Chair in charge
Nobuko Matsushima (Janssen Pharma K.K.) Leader
Risako Sekiguchi (Novartis Pharma K.K.)
Mika Tsujimoto (Eli Lilly Japan K.K.)
Tomoyuki Wada (Senju Pharmaceutical Co., Ltd.)
Takehiko Hiraoka (Maruishi Pharmaceutical Co., Ltd.)

Acknowledgments: The authors would like to thank Takashi Asakawa of Chugai Pharmaceutical, Vice-Chairman of the Drug Evaluation Committee of the Japan Pharmaceutical Manufacturers Association, and the regular vice-chairmen and auditors of the Data Science Subcommittee for their peer review and helpful advice in preparing this report.

Column: Effect Modifiers and Related Terms

While we are trying to use the term of "effect modifiers" as much as possible in this report, some sections use other terms which are commonly used in a particular application. Those terms are summarized here.

Effect modifiers

Factors which show evidence of that they affect (in other word, modify) the treatment effect (efficacy and safety) of a particular drug.

Risk factors

Factors which have a potential to increase a risk of causing adverse reactions to a drug. Risk factors are effect modifiers in the safety framework.

Predictive factors

Factors used to identify individuals who are more likely to experience favorable or unfavorable effects from exposure to a drug than individuals who do not have the factors.[21] The concept of effect modifiers is much broader, and predictive factors are part of them. They may be used for enrichment purposes to select/limit the target population for a clinical trial when the presence or absence of a treatment effect is already clearly known.

Prognostic factors

Factors used to identify the likelihood of clinical events, disease recurrence or progression in patients with the disease or condition of interest, regardless of existence of treatment[21]. Prognostic factors can be a part of effect modifiers.

Background/Demographic factors

Baseline (pre-treatment) disease-related factors (including prognostic factors) and individual characteristics such as age, gender, country of participation, and race of subjects participating in a clinical trial or an observational study. Information necessary to characterize the population of the clinical trial or observational study.

---

[21] FDA-NIH Biomarker Working Group, BEST (Biomarkers, Endpoints, and other Tools) Resource, Last Updated: November 29, 2021 (https://www.ncbi.nlm.nih.gov/books/nbk326791/pdf/Bookshelf_NBK326791.pdf; last access confirmed on July 21, 2022)

Covariates

In a statistical model, explanatory variables which affect the response variable. A change in a covariate is expected to result in a corresponding change in the response variable. In the narrow meaning (especially in statistical analysis software), they are sometimes considered continuous variables out of explanatory variables, but are not limited to this definition.

Stratification factors

For a clinical trial mainly conducted with a randomized controlled design, balancing the distribution of factors, which are already known to affect treatment effects, among the comparison groups is critical to obtain unbiased study results. The operation during randomization is called stratified allocation. The factors considered at that time are called stratification factors. For examples, prognostic factors of the disease, the regions, and the study sites are used as stratification factors.

(Back to the main text) (Back to the table of contents)

Column: Genetic Polymorphisms of Metabolic Enzymes and Transporters

Some metabolic enzymes and transporters have genetic polymorphisms that cause changes in their expression levels and/or functional activity. This is known to lead to interindividual differences in drug exposure when a certain drug is administered, and ultimately to interindividual differences in therapeutic outcome (efficacy and safety). It is also known that racial/ethnic differences in the frequency of some genetic polymorphisms are observed. For example, there are many reports on genetic polymorphisms of cytochrome P450 (CYP) and UDP-glucuronyltransferase (UGT) or of transporters, which are involved in the metabolism and disposition of many small molecule drugs. There have been multiple reports comparing genetic polymorphism frequencies within Asians, such as Japanese and Chinese, or with other races.[22,23] In 2016, a study adopted by the Ministry of Health, Labour and Welfare (MHLW) to promote the practical application of innovative drugs, medical devices and regenerative medicine products, "Scientific Information for the Evaluation of Genetic Polymorphisms in Nonclinical and Clinical Phase I Studies of Pharmaceuticals (Final Draft)," was published, which compiled methodology on the types of genetic polymorphisms that may be considered in Phase I clinical trials in Japanese subjects and how to evaluate their effects. [24]

Below are examples of genetic polymorphisms of metabolic enzymes and transporters and their effects on activity, as well as differences in gene polymorphism frequencies between races.

CYP2C19, which is involved in the metabolism of many drugs such as proton pump inhibitors, is known to have two polymorphisms, known as *2 (681G>A) and *3 (636G>A), that cause deficiency in activity, and there are large racial differences in their frequency[23]. The percentage of individuals homozygous or heterozygous for *2 and *3 (*2/*2, *3/*3, *2/*3),

---

[22] Michael Man et al., Genetic Variation in Metabolizing Enzyme and Transporter Genes: Comprehensive Assessment in Three Major East Asian Subpopulations With Comparison to Caucasians and Africans. *Journal of Clinical Pharmacology* **50(8)** 929-940 (2010).

[23] Kurose Kouichi et al., Population differences in major functional polymorphisms of pharmacokinetics/pharmacodynamics-related genes in Eastern Asians and Europeans: implications in the clinical trials for novel drug development. *Drug Metabolism and Pharmacokinetics* **27(1)** 9-54 (2012).

[24] Graduate School of Pharmaceutical Sciences, Tohoku University. (2018). Scientific Information for the Evaluation of Genetic Polymorphisms in Nonclinical and Clinical Phase I Studies of Pharmaceuticals (Final Draft). (https://www.pmda.go.jp/files/000221578.pdf; last access confirmed May 24, 2022)

which represent poor metabolizers with little CYP2C19 activity as a phenotype, reaches about 20% in Japanese, but only a few percent in Caucasians. The majority of Caucasians have been reported to be normally active, extensive metabolizers.[22]

UGT1A1, one of the molecular species of UGT, the frequency of *6 (211G>A, G71R), a mutation causing reduced activity, is reported to be about 18% in Asians, whereas it is less than 1% in Caucasians. On the other hand, the mutation causing down-regulation, *28 (polymorphism in the number of TA repeats in the gene promoter region), is reported to have a frequency of about 10% in East Asians, compared to 30% to 40% in Caucasians and Blacks[23] Since UGT1A1 is involved in the metabolism of the active metabolite of the anticancer drug irinotecan, studies were conducted to link its decreased activity to adverse effects (neutropenia) caused by irinotecan administration. As a result, in 2005, the package insert for CAMPTOSAR® in the US was revised to state, "a reduction in the starting dose by at least one level of CAMPTOSAR should be considered for patients known to be homozygous for the UGT1A1*28 allele."[25] In Japan, the package insert was revised in 2006 to state that "With regard to two genetic polymorphisms (UGT1A1*6 and UGT1A1*28) of UDP-glucuronosyltransferase (UGT), the main metabolizing enzyme of the active metabolite (SN-38) of this drug, the possibility of serious side effects (especially neutropenia) has been reported to increase due to decreased glucuronidation activity of UGT1A1 and delayed metabolism of SN-38 in patients homozygous for either (UGT1A1*6/*6, UGT1A1*28/*28) or heterozygous for either (UGT1A1*6/*28)."[26] Both *6 and *28 are genetic mutations that cause decreased activity in glucuronide conjugation, but the frequency of expression of the genetic mutation differs by race, and this is an example of the difference in the description of the drug on the package insert between countries.

Polymorphism of organic anion transporting polypeptide 1B1 (OATP1B1), is a well-known example of the gene mutation that alters the activity of the transporter. OATP1B1 is expressed on the sinusoidal membrane of hepatocytes and involved in the hepatic uptake of many drugs such as HMG-CoA reductase inhibitors. Several genetic polymorphisms have been reported for the gene encoding OATP1B1, the solute carrier organic anion transporter family member 1B1 (SLCO1B1). For example, the polymorphism 521T>C (V174A) is known to cause

[25] CAMPTOSAR label: Drugs@FDA: FDA-Approved Drugs: (https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm?event=overview.process& (ApplNo=020571; last access confirmed May 24, 2022)

[26] Irinotecan Hydrochloride Attachment. (https://www.info.pmda.go.jp/psearch/PackinsSearch?dragname=%A5%A4%A5%EA%A5%CE%A5%C6%A5%AB%A5%F3%B1%F6%BB%C0%B1%F6%BF%E5%CF%C2%CA% AA; last access confirmed May 24, 2022)

decreased expression and reduced transport activity, and its frequency is low in Blacks and about 15% in East Asians and Whites.[23] On the other hand, 388A>G (N130D), which is found at an allele frequency of about 60% in Japanese, has a tendency to slightly increase activity. 521T>C (*5, *15, and *17) generally shows the decrease the hepatic uptake activity of the substrate drug, resulting in increased blood concentration of the drug. On the other hand, it has been reported that 388A>G tended to decrease the blood concentration of the substrate drug.[27]

The results of a study of the genotypes of major CYPs and phenotypes with typical substrate drugs in more than 600 Japanese (first or third generation Japanese living in Japan and overseas), Korean, Chinese, and Caucasian subjects were reported[28].

Based on the mean metabolite ratios of typical substrate drugs, the metabolic activity (phenotype) of key CYPs was assessed and found to be similar among each race, except that Caucasians tend to show lower CYP2D6 activity and Asians tend to show lower CYP2C19 activity. For example, the metabolic activity of CYP2C19 showed that the *2 and *3 alleles resulted in reduced metabolic activity regardless of race, and the *2/*2, *2/*3 and *3/*3 homozygous or heterozygous phenotypes showed a low metabolic type (poor metabolizer) regardless of race.

In the examples presented above, genetic polymorphisms can be patient factors (intrinsic factors) that influence pharmacokinetics and, ultimately, efficacy and safety. When comparing pharmacokinetics, efficacy, and safety in a specific racial group or country/region, differences may be observed due to differences in gene polymorphism frequencies among racial groups, even though no differences are observed in populations with the same genotype. In this case, it is not the factor of race that affects pharmacokinetics, efficacy, or safety, but the factor of genetic polymorphism.

(Back to the main text) (Back to the table of contents)

[27] Maeda Kazuya, Organic Anion Transporting Polypeptide (OATP)1B1 and OATP1B3 as Important Regulators of the Pharmacokinetics of Substrate Drugs. *Biological & Pharmaceutical Bulletin*, **38(2)**155-168 (2015).

[28] Myrand, S.P. et al. Pharmacokinetics/genotype associations for major cytochrome P450 enzymes in native and first- and third-generation Japanese *Clinical Pharmacology & Therapeutics*, **84(3)** 347-361 (2008).

Column: Exploring Effect Modifiers -Alogliptin Example

Alogliptin (Nesina tablets) is one of the cases in which intrinsic and extrinsic ethnic factors in a drug response were examined.

Alogliptin is a dipeptidyl peptidase-4 (DPP-4) inhibitor that delays inactivation of the incretin hormones, glucagon-like peptide-1 (GLP-1) and glucose-dependent insulinotropic polypeptide (GIP). In Japan, the drug was approved in April 2010 for the treatment of type 2 diabetes patients who fail to respond adequately to diet and exercise alone or diet and exercise plus an alpha-glucosidase inhibitor. For the submission of the application, the monotherapy was developed based on a bridging concept, while the combination therapy with an α-glucosidase inhibitor was locally developed in Japan. For monotherapy, the local Phase II study was positioned as a bridging study, and the foreign Phase II (dose-response study) and Phase III studies were included in the clinical package. [29]

In the application, the similarities and differences between Japan and the West (Europe and the US) in intrinsic and extrinsic ethnic factors in type 2 diabetes were compared. Based on the available findings and external data, ethnic differences in insulin secretion capacity, insulin resistance, and BMI were examined, and it was stated that there are ethnic differences in these factors based on external data. As for extrinsic ethnic factors, there are no major differences in diabetes diagnostic criteria and treatment goals of pharmacotherapy between Japan and the West, but there may be differences in the ratio of dietary nutrients.

In the clinical development of alogliptin, the dose-response relationship in patients was examined with the change in HbA1c. Although the HbA1c changes in the bridging study in Japanese patients were greater than those in studies in non-Japanese patients, the dose-response relationship was considered to be similar among the three studies, since the HbA1c reduction in the 6.25-mg group was smaller than that in the ≥12.5-mg groups in all studies and the ≥12.5-mg groups showed similar HbA1c reductions among all studies. In addition, BMI, C-peptide level at fasting state, HbA1c baseline, and insulin resistance were examined

[29] Nesina Tablets Review Report (2010). (https://www.pmda.go.jp/drugs/2010/P201000029/400256000_22200AMX00309_A100_3.pdf; last access confirmed May 24, 2022)

as factors that may influence the efficacy of alogliptin, which are thought to influence the etiology and pathophysiology of type 2 diabetes. Subgroup analyses showed reduction in HbA1c in the groups receiving alogliptin or alogliptin ≥12.5 mg compared to the placebo group in each stratum in all studies.

In the review by PMDA, the extrapolation of the foreign studies to Japanese population is considered possible, because although it cannot be said that the similarity of the dose-response relationship was clearly demonstrated, the mechanism of action is enzyme inhibition, and the HbA1c reductions in the groups receiving 12.5 mg or more were similar in the Japanese study and the foreign studies. The results of subgroup analyses by background factors showed that HbA1c-lowering effect tended to improve in the 12.5-mg or more dose group compared to the placebo group in each stratum. Although subgroup analyses stratified by each background factor were performed, no conclusions were drawn regarding the impact and extent of efficacy of each factor.

Subsequently, a number of meta-analyses examining the effect of race and other factors on the efficacy of alogliptin or DPP-4 inhibitors have been published, with several papers showing greater efficacy in Asians compared to non-Asians.[30,31,32] For example, a meta-analysis of 15 randomized clinical trials of alogliptin showed that the HbA1c-lowering effect of alogliptin in patients with type 2 diabetes and the percentage achieving target HbA1c were higher in Asians than in non-Asians.[32] In this study, a meta-analysis of randomized trials comparing alogliptin with placebo as monotherapy or add-on therapy in adult patients with type 2 diabetes was conducted, with subgroups of Asian-majority trials and non-Asian majority trials. Results showed that the reduction in HbA1c was greater in the Asian-majority trial (−0.75% [95% CI −0.84:−0.65]) than in the non-Asian-majority trial (−0.61% [95% CI −0.68:−0.54]). Efficacy tended to be higher in the Asian-majority trials in terms of percentage of patients achieving target HbA1c and postprandial hypoglycemic effect.

DPP-4 is highly expressed in visceral fat of obese individuals and is transferred to the systemic circulation. DPP-4 activity is also higher in obese individuals. Therefore, the efficacy of DPP-4 inhibitors may be lower in non-Asian patients with high BMI (a measure of the degree of

[30] Kim, G.Y. et al. Differences in the glucose-lowering efficacy of dipeptidyl peptidase-4 inhibitors between Asians and non-Asians: a systematic review *Diabetologia* **56(4)** 696-708 (2013).
[31] Fujita, K. et al. Factors Related to the Glucose-Lowering Efficacy of Dipeptidyl Peptidase-4 Inhibitors: A Systematic Review and Meta-Analysis Focusing on Ethnicity and Study Regions. *Clin Drug Investig* **37(3)** 219-232 (2017).
[32] Cai, Y. et al. Ethnic Differences in the Efficacy and Safety of Alogliptin: A Systematic Review and Meta-Analysis. *Diabetes Ther.* **9(1)** 177-191 (2018).

obesity) than in Asian patients. In fact, clinical trials of other DPP-4 inhibitors and meta-analyses have shown a significant correlation between baseline BMI and HbA1c lowering effect[30, 33] The racial differences in the efficacy shown in the meta-analyses might be explained by differences in insulin sensitivity and differences in the frequency of genetic polymorphisms affecting insulin sensitivity. It may also be caused by difference in diet, since the reduction of HbA1c by DPP-4 inhibitors is significantly correlated with the estimated intake of fish, EPA and DHA, and the serum concentrations of EPA and DHA.[30, 31, 32] Some of these factors are still only possibilities.

As in the alogliptin example above, intrinsic and extrinsic ethnic factors in drug efficacy or safety cannot be concluded from a single study of a single drug. On the other hand, it may be possible to examine effect modifiers early in the development of a drug based on its mechanism of action, as the relationship between DPP-4 activity and obesity, the drug's target, was observed.

(Back to the main text) (Back to the table of contents)

[33] Yagi, S. et al. Predictive Factors for the Efficacy of Dipeptidyl Peptidase-4 Inhibitors in Patients with Type 2 Diabetes Mellitus. *Diabetes Metab J.* **39(4)**. 342-347(2015).

Column: Exploring Effect Modifiers: Gefitinib Example

Gefitinib (brand name Iressa) may have triggered the focus on effect modifiers in drug development.[34] This drug was subsequently approved for the treatment of advanced non-small cell lung cancer (NSCLC) in Japan in 2002 and conditionally approved in US in 2003 based on the results of Phase II clinical trials for monotherapy with antitumor activity as the primary endpoint (MRCT conducted mainly in Japan and Europe [IDEAL1], and a foreign trial conducted in the United States [IDEAL2]). In IDEAL1, a higher response rate was observed in the Japanese subgroup compared to the non-Japanese subgroup (27.5% vs. 9.6%) and a logistic regression analysis was conducted to investigate patient background factors relevant to efficacy. The results showed that factors such as performance status and histological type contributed more than ethnicity (Japanese vs. non-Japanese), and it was thought the observed inconsistency was due to differences in the distribution of these factors between the Japanese subgroup and the non-Japanese subgroup.[35] Foreign Phase III trials conducted outside Japan showed that gefitinib in combined platinum-based chemotherapy showed no additional efficacy in chemotherapy-naïve patients with advanced NSCLC, and therefore subsequent confirmatory trials were conducted to evaluate gefitinib monotherapy.

Other foreign phase III trials (ISEL and INTEREST), which were required as a condition for FDA approval, were conducted for patients with locally advanced or metastatic NSCLC who had received one or two previous chemotherapy regimens, with overall survival (OS) as the primary endpoint. The ISEL was to show the superiority of gefitinib over best supportive care, and the INTEREST was to show the non-inferiority of gefitinib to docetaxel. The results of ISEL were reported in 2004, and there was a statistically significant improvement in anti-tumor activity, but the superiority of gefitinib over best supportive care in the primary endpoint of OS was not shown.[36] The results led to restrictions on the use of the drug in the US in 2005, and the sponsor withdrew marketing authorization application, which was submitted in Europe in 2003. However, subgroup analyses in ISEL suggested that gefitinib might be effective in patients with no history of smoking and patients of Asian origin, and subsequent

[34]Armour, A.A. and Watkins, C.L. The challenge of targeting EGFR: experience with gefitinib in nonsmall cell lung cancer. *Eur Respir*. **19(117)** 186-96 (2010).

[35] Examination Report. (https://www.pmda.go.jp/drugs/2002/P200200028/67022700_21400AMY00188_110_2.pdf; last access confirmed May 25, 2022)

[36]AstraZeneca Corporation Press Release, Iressa in Advanced Non-Small Cell Lung Cancer ISEL Study Results (https://www.astrazeneca.co.jp/media/press-releases1/2004/20041220.html#; last accessed (confirmed May 25, 2022)

investigations showed that gefitinib tended to work better in patients of Asian origin, female sex, no history of smoking, and adenocarcinoma histology. Consequently, phase III trial was conducted in the patients selected based on these clinical characteristics (no history of smoking or former light-smokers who had adenocarcinoma of NSCLC and no prior chemotherapy, conducted in Asia) to compare the primary endpoint of progression-free survival (PFS) between gefitinib and carboplatin/paclitaxel combination therapy (IPASS). In 2008, the results of IPASS showed the non-inferiority and superiority of gefitinib to the carboplatin/paclitaxel combination statistically, but the treatment effect was not constant over time. The Kaplan-Meier curves of PFS were crossed in the middle of the study. The preplanned analysis based on the biomarker status showed that the presence of the EGFR mutation was a very strong predictor (effect modifier). In 2007 the results in INTEREST showed the inferiority of gefitinib to docetaxel. In 2009, marketing authorization was granted in Europe for gefitinib for adults with locally advanced or metastatic NSCLC with activating mutations of EGFR-tyrosine kinase, regardless of prior therapy.[37] In 2011, the indication was also changed to inoperable or recurrent NSCLC with EGFR mutation-positive disease in Japan.

As described above, gefitinib single-agent was approved in Japan and the US based on the results of antitumor effect in phase II trials, but the application was withdrawn in Europe because the cause of the ethnic difference observed in the trials could not be identified, and additional efficacy over chemotherapy and life extension in a single-agent setting were not as expected in the subsequent evaluations. However, further investigations have identified the clinical characteristics benefiting from gefitinib more, and finally identified the true predictive factor, the presence of an EGFR mutation underlying these clinical characteristics. Although it took many years and extensive research to identify the appropriate target patient population for treatment, in retrospect, this can be regarded as an example of specific processes of encounter with a considerable ethnic difference through simultaneous global development involving multiple regions, exploration of the causes, success in identifying the effect modifiers, and leading to appropriate use of a drug.

(Back to the main text) (Back to the table of contents)

[37] AstraZeneca Corporation Press Release, Iressa® Receives Marketing Authorization Recommendation in Europe (https://www.astrazeneca.co.jp/media/press-releases1/2009/2009050702.html#; last access confirmed May 25, 2022))

Column: What is a Region and What is Consistent Evaluation of Treatment Effects Across Regions?

In 2018, the ICH E17 Guideline was issued, the first guideline for aiming submission across multiple regions simultaneously. Before that, we conducted many MRCTs and did submissions. Regardless of whether before or after the E17 Guideline, it is thought that the evaluation of the consistency of treatment effects across regions in MRCTs and the estimation of the regional treatment effect are important information for regulatory authorities to decide whether to approve a drug for marketing authorization in their region. A region which is subject to estimate the regional treatment effects is the one that the authorities are considered to be in charge of, and it may be a single country (e.g., MHLW/PMDA for Japan) or pooling of several countries (e.g., EMA for EU member countries). As the number of countries participating in MRCT increases and the sample size per country is decreases accordingly, there are increasing concerns about the reliability of treatment effect estimates from subjects in a single country. This has become a particularly serious issue when the authority is charge of a single country. We have considered the possibilities and limitations of pooling geographically close countries (usage of region) as one approach to address this issue. How will the evaluation of the consistency of treatment effects across regions and evaluation of the regional treatment effect change under the E17 Guideline?

[Definition of the three types of regions in the E17 Guideline]

Generally, a region is a set of countries which are geographically close and contiguous each other, and in the context it is often intended to have a common uniqueness or some equitable feature that allows it to define its boundaries with other regions on that basis. Thus, if we consider a region as a set of countries that have some commonality, the three types of regions described in the E17 Guideline (geographical regions, regulatory regions, and pooled regions) may be thought of as follows.

Region	Definition in the E17 Guideline	Commonality focused to pool countries
Geographic region*		Geographic location
Regulatory region	A region comprised of countries for which a common set of regulatory requirements applies for drug approval (e.g., EU).	Regulatory Requirements

Pooled regions**	Pooling some geographical regions, countries or regulatory regions 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.	Important ethnic factors affecting treatment effect and their distribution (e.g., access to standard of care, severity of illness affecting treatment effects)

* "a geographical region" in the definition of region "Regulatory Region:
A region comprised of countries for which a common set of regulatory requirements applies for drug approval" in the 3. Glossary of the E17 Guideline.

** an organized approach in the E17 Guideline

The pooled regions, a pooling strategy focusing on commonality of intrinsic and/or extrinsic factors affecting the treatment effect (that is, effect modifiers), have the following features: To define it, the intrinsic and/or extrinsic factors affecting the treatment effect and the disease under investigation should be identified, and there is enough evidence. If the health authorities agree on that, the pooled regions can be prespecified and used for planning MRCTs.

The countries/regions to be pooled do not necessarily have to be geographically contiguous; and distant countries/regions such as enclaves may be pooled. Previously, we sometimes wondered if the estimates of the treatment effect obtained from data in East Asia could be used as a reference for evaluation of the treatment effect in Japanese subjects when the number of Japanese subjects participating in MRCTs was extremely small and it was difficult to rely on the estimate of the treatment effect obtained from the Japanese subjects. Behind these thoughts, we had expectation that East Asia, pooling contiguous countries to Japan (geographic region) can be a reference for evaluation of the treatment effect in Japan because contiguous countries to Japan should have similar intrinsic and/or extrinsic factors and then the treatment effect on these countries should be similar. Pooled regions and geographic region have the same intention, that is to pool countries/regions with the similar distributions of intrinsic and/or extrinsic factors affecting treatment effects, and it can be said that both aim in the same direction. Pooled regions differ from geographic regions in that it required evidence of the commonality in intrinsic and/or extrinsic factors affecting the treatment effect and the agreement from the authority on that, and distant countries/regions can be pooled. For a pooled region, the treatment effects among countries within the region are expected to be similar because the countries are pooled based on the evidence of the intrinsic and/or extrinsic factors affecting the treatment effect and the disease under

investigation, and the treatment effect in the region can be more representative of a country within the region than the geographic region can.

[Consistency of treatment effects across regions].

Before the E17 Guideline was issued, we evaluated the consistency of the treatment effect across regions based on regions with pooled countries which are geographically close to each other (e.g., Asia, Europe), hoping that the treatment effects would be consistent across regions. If clinically relevant differences have been observed, we conducted exploratory analyses to look for any factors that contributed to the differences (with a sort of culprit-seeking mindset). When such exploratory analyses led to a discovery of intrinsic and/or extrinsic factors which potentially affected the treatment effect and a possibility that different distributions of the identified factors across regions caused a difference of treatment effect across regions, we reevaluated the regional consistency by adjusting impacts of the intrinsic and/or extrinsic factors, but did not evaluate the regional consistency of the treatment effect reflecting any potential impact of the intrinsic and/or extrinsic factors.

What is the assessment of the consistency across regions under the E17 Guideline?

E17 Training material explains that "If clinically relevant differences among regions are observed, then the MRCT provides a unique opportunity for additional learning about the factors that may explain these differences (the differences of the treatment effect across regions)"[38] and sees the evaluation of the regional consistency as the beginning of new learning to better understand the treatment effect. In addition, we can discuss the regional treatment effect reflecting the impact of the intrinsic and/or extrinsic factors affecting the treatment effect using pooled regions (see the note below). Thus, the evaluation of the regional consistency based on pooled regions would be to confirm the regional inconsistency of the treatment effect estimated based on the evidence.

> (Note: Consider pooled region under a hypothetical scenario in which diet is known to be a factor affecting the treatment effect for type 2 diabetes. For example, it is expected that high-fat and/or low-carbohydrate content of meals lead to higher treatment effect, and low-fat and/or high-carbohydrate of content of meals to lower treatment effect in this compound (that is, there is a quantitative interaction between treatment and diet). If the evidence is sufficient and the authorities agree, pooled regions can be prespecified

[38] Training Materials on "General Principles for the Planning and Design of Global Clinical Trials" (https://www.pmda.go.jp/int-activities/int-harmony/ich/0022.html; last access confirmed May 25, 2022)

as follows and regional allocation of sample size and the evaluation of the regional consistency can be done based on pooled regions.

Region A: pooling countries where high-fat and/or low-carbohydrate content of meals are common/popular

Region B: pooling countries where high-fat and/or low-carbohydrate content of meals are common/popular

The distributions of diet are similar among countries within region and then the treatment effects of countries within the region can be expected to be similar. It can be guessed that the treatment effect in Region A will be higher than one in Region B (there is a quantitative interaction between treatment and diet) in advance.

As described above, future evaluations of the regional consistency of treatment effects would not focus only on concluding that the treatment effects are consistent across regions, but also require to learn how the treatment effects are affected by the intrinsic and/or extrinsic factors and evaluate whether the results on the entire population of MRCT can be applied to all regions. In particular, an evaluation based on pooled regions would provide an opportunity to understand the treatment effects from a different perspective. Pooled regions is pooling of countries with similar distributions of the intrinsic and/or extrinsic factors affecting the treatment effect and the disease under investigation, and the estimates of the regional treatment effects from pooled regions reflect the impact of these factors. Thus, it is estimated in advance there are differences in the treatment effects among pooled regions. In other words, the evaluation of the consistency of treatment effects based on pooled regions is to evaluate whether the differences anticipated based on the evidence are actually observed. When an authority is in charge of a single country, the sample size of the country in the MRCT may be too small to convince themselves with the estimate of the treatment effect from the data of the country. Because the estimates of treatment effects based on pooled regions reflect the impact of the ethnic factors which are part of the country, they may provide information to aid in benefit-risk evaluations for the country that comprise the pooled region.

In practice, there may be many cases where there is insufficient evidence on the intrinsic/extrinsic factors related to a targeted disease or an investigational drug to define pooled regions at the MRCT planning stage. In these circumstances, when unanticipated regional differences are observed, or when candidates for the intrinsic/extrinsic factors related to the target disease or the investigational drug are identified, the evaluation of treatment effect based on pooled regions defined in an exploratory manner may provide a

deeper understanding of how much treatment effects are influenced by the intrinsic/extrinsic factors, and this understanding may be used in designing and planning of future clinical trials. This strategy of pooled regions could also be adapted to evaluate not only efficacy but also safety.

(Back to the main text) (Back to the table of contents)

Column: Using Effect Modifiers for Prediction -Toward Personalized Therapy

The following model is illustrated in the main text.

Y = f(treatment, weight, biomarker A, genetic information K) + error

Here, the model is formulated with treatment and identified effect modifiers, which are weight, biomarker A, and genetic information K, as the explanatory variable and response to treatment (e.g., efficacy and safety events) Y as the response variable. The term f (treatment, weight, biomarker A, genetic information K) gives the average value for individuals with a particular combination of the three effect modifiers (weight, biomarker A, and genetic information K) who received the treatment. The error represents the remaining variation that is not accounted for by the treatment and effect modifiers (weight, biomarker A, and genetic information K).

Conversely, when assuming a response Y to be obtained or avoided, the model may be useful in determining the optimal treatment if information on effect modifiers, namely body weight, biomarker A, and genetic information K, is available.

Model-Informed Precision Dosing (MIPD) utilizes modeling & simulation to predict the treatment most likely to achieve a favorable benefit-risk balance based on individual patient characteristics. The basic form of MIPD is to predict the appropriate dose most likely to achieve a favorable benefit-risk balance based on individual patient characteristics. There is growing interest in MIPD in drug dosing decisions, particularly for drugs with narrow therapeutic windows, drugs with high variability in response among patients, and drugs for special populations such as pediatrics, patients with significant organ dysfunction, and the elderly.[39, 40, 41] For example, there is a report on a randomized clinical trial of paclitaxel in Chinese patients with non-small cell lung cancer, comparing standard dose per body surface area to a model-based adjusted dose.[42] Because the target time to maintain plasma

[39] Thomas M. Polasek et al. Toward Dynamic Prescribing Information: Codevelopment of Companion Model-Informed Precision Dosing Tools in Drug Development. *Clin Pharmacol Drug Dev* **8**(4) 418-425 (2019).

[40] Adam S. Darwich, et al. Model-Informed Precision Dosing: Background, Requirements, Validation, Implementation, and Forward Trajectory of *Annu. Rev. Pharmacol. Toxicol* **61**:225-245(2021).

[41] Daniel F.B. Wright et al. Spotlight Commentary: model-informed precision dosing must demonstrate improved patient outcomes. *Br. J. Clin. Clin. Pharmacol.* **85**:2238-40 (2019).

[42] Zhang, J et al. Randomized study of individualized pharmacokinetically-guided dosing of paclitaxel compared to body-surface area dosing in Chinese patients with advanced non-small cell lung cancer. *Br J Clin Pharmacol* **85**(10):2292-2301(2019).

paclitaxel concentration higher than 0.05 μmol/L is 26-31 hours, the adjusted-dose group of the study used plasma paclitaxel concentration and pharmacokinetic modeling to individually predict the time for paclitaxel concentration to exceed 0.05 μmol/L. In addition, dose adjustments were made based on the presence or absence of neutropenia. As a result, Grade 4 hematologic toxicity and neutropenia were significantly reduced in the adjusted dose group, with hazard ratios of 0.59 (95% CI 0.39-0.87) and 0.57 (95% CI 0.37-0.83), respectively. The results showed that efficacy endpoints such as objective response and disease control did not differ between study groups, while progression-free survival was longer in the adjusted dose group. In other words, individualized dosing reduced safety risk without a reduction in efficacy.

Modeling and simulation techniques have evolved and are now an integral part of most drug development programs. Modeling and simulation are routinely used to clarify the relationship between drug dosing (exposure) and response (efficacy and safety), to support the dosing rules in the package insert (adding insight to existing information), and to inform clinical trial design, including setting the dose (helping to predict the future). On the other hand, its use in clinical practice is still limited. MIPD is expected to provide a means to predict drug response and dosing in individual patients by quantifying and explaining factors of inter and intrapatient variability, thereby providing the most appropriate treatment for each patient.

Inter and intrapatient variability in drug response involves multiple factors. In addition, the understanding of effect modifiers for drug response is often less complete than the understanding of effect modifiers for pharmacokinetics, and better predictive models may be constructed by pooling response data for multiple drugs. Accurate models and an environment which enables its utilization in clinical practice are essential for the implementation of MIPD. The accumulation of data on a wide range of patient factors, environmental factors, and drug responses in actual clinical practice is expected to further improve the accuracy of the model.

(Back to the main text) (Back to the table of contents)

Column: History of Drug Development: From Local Development to Global Simultaneous Development

Movement from local drug development to international standards

A clinical development framework in which a clinical trial is conducted in multiple countries/regions from the world under the same protocol and the data from such trials are submitted to multiple regulatory authorities for drug approval is probably not uncommon nowadays. Perhaps there are many readers who are familiar with such a clinical development strategy only?

Until the mid-1990s, each country had its own regulatory requirements for new drug approval. Under those regulatory requirements, pharmaceutical companies conducted clinical trials and executed drug development in each country by taking enormous effort, expense, and resources for a long period of time. Pharmaceutical companies often started drug development first in their home country, obtained approval there, and then moved to other countries under their regulatory requirements. Although experiences of drug development and finding through drug development program in their home country could be utilized for that in subsequent countries, they still had to conduct many duplicate clinical trials.

In 1990, against a background of concern over the growing cost of drug development, 6 from regulatory authorities and industry associations of 3 regions, Europe, the United States, and Japan, launched the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH)[43] (see note below). With the common goal of delivering better drugs to patients around the world faster, activities were undertaken to harmonize international standards for the review of new drug approvals of the regulatory authorities in each region, and the differences in regulatory requirements were greatly improved. Since GCP (Good Clinical Practice), the practice standard for clinical trials was agreed to in 1996, common scientific and ethical standards were established, and it laid the foundation for utilizing foreign clinical trial data that is in place today. The important guidelines leading to standardization of not only regulatory requirements but also a common set of concepts and reporting formats throughout the world, e.g. clinical safety reporting (ICH

[43] Pharmaceuticals and Medical Devises Agency (PMDA), ICH International Conference on Harmonization of Pharmaceutical Regulations (https://www.pmda.go.jp/int-activities/int-harmony/ich/0014.html; last accessed (confirmed May 25, 2022)

E2), the Medical Dictionary for Regulatory Activities (ICH M1; MedDRA), the guideline for the structure and content of clinical study reports (ICH E3), the general considerations for clinical trials throughout the process of clinical development (ICH E8), the statistical principles for clinical trials (ICH E9), and the Common Technical Document (M4) were agreed upon in the mid-to-late 1990s.

(Note) In 1990, the EC member countries in Europe were conducting regulatory reviews separately, but the European Agency for the Evaluation of Medicinal Products (EMEA), the predecessor of the European Medicines Agency (EMA), was established in 1995 after seven years of discussions, and created a framework for mutual acceptance of data among EU member countries. It could be said that overlap between the establishment of the EMEA and the beginning of international regulatory harmonization after the ICH established, was not coincidence but inevitable. Refer to Doi[44] for more background on the early days of the ICH.

Bridging Strategy: Acceptance of Foreign Clinical Trial Data

As the global standardization and streamlining of drug development progressed, clinical trial data conducted in foreign countries has been used with hesitation because of differences in attitude and standards on clinical trial quality and data reliability, as well as ethnic and regional differences. However, if duplication of clinical trials at a global level can be minimized and clinical trial data can be used across countries and regions, the benefits in terms of saving human, economic, and time resources will be great. At this request for the expansion of use of foreign clinical trial data in applications for drug approval, the issues of ethnic difference and regional difference were scientifically discussed, and the development of guidelines for accepting foreign clinical trial data in Japan, US, and EU was proposed[45].

In 1998, the ICH E5 guideline "Ethnic Factors in the Acceptability of Foreign Clinical Data" was issued.[46] The E5 Guideline defined ethnic factors as the genetic and physiological (intrinsic)

---

[44] Osamu Doi, Pharmaceutical Regulatory Harmonization International Conference (ICH), Pharmaceutical and Medical Device Regulatory Science 41(8) 636-637 (2010) (https://www.pmrj.jp/publications/02/pmdrs_column/pmdrs_column_08-41_08.pdf; last access confirmed May 25, 2022)

[45] Hajime Yasuhara, Overview of ICH E5, Clinical Pharmacology 32(4) 143-144 (2001) (https://www.jstage.jst.go.jp/article/jscpt1970/32/4/32_4_143/_pdf; last access checked May 25, 2022)

[46] Ethnic Factors to Consider When Accepting Foreign Clinical Data, Pharmaceutical Affairs

characteristics and cultural and environmental (extrinsic) characteristics of a population, and described the basic concept to evaluate the influence of ethnic factors appropriately, the concept to determine whether clinical study results of drugs that had been shown to be effective in foreign countries can be extrapolated (bridged) to the population of a new region, and the requirements for a bridging study (clinical trial conducted in a new region to extrapolate foreign clinical data to the new region). In Japan, new drugs or new indications have been approved based on the bridging strategy since around 2000, and, more than half of drug development projects were based on the bridging strategy around 2010[47]. An important point in determining whether foreign clinical data can be extrapolated to a new region is whether dose-response, safety, and efficacy are similar between foreigners and Japanese. Specifically, results were presented on the similarities or differences between foreign clinical trial data and bridging study data conducted in Japan, and the ability to extrapolate foreign clinical data was evaluated in terms of similarity between intrinsic and extrinsic factors in applications for approval based on the bridging strategy.

On the other hand, around 2005, the media began to focus on "drug lag," drugs which have already been available in foreign countries are not available in Japan or have taken a long time to become available in Japan, and this was recognized as a serious problem.[48] Although the bridging strategy made it possible to shorten the time to application by extrapolating data from confirmatory studies in foreign countries and skipping confirmatory studies in Japan, in reality, pharmacokinetic studies and dose-response studies in Japanese subjects (conducted as bridging studies) were often required, and drug lag remained due to the delayed start of these studies. In order to resolve this problem, recognition was fostered that it was important to align the start of clinical development required for regulatory approval in Japan with the start of clinical development in other countries.

Council No. 672, August 11, 1998 (https://www.pmda.go.jp/files/000156571.pdf; last access confirmed May 25, 2022)

[47] Hiyoyuki, Uesaka, <Review Article>Simultaneous Global Development of Pharmaceuticals and Multiregional Trials, Health & Medical Sciences 60(1) 18-26 (2011) (in Japanese). (https://www.niph.go.jp/journal/data/60-1/201160010005.pdf; last access confirmed May 25, 2022)

[48] Taku Serio, Drug Development in Japan from the Perspective of Foreign Pharmaceutical Companies, Clinical Hematology 50(7) 556-562 (2011). (https://japhmed.jp/%e8%87%a8%e5%ba%8a%e8%a1%80%e6%b6%b2%e8%aa%8c20090421.pdf; last access confirmed May 25, 2022)

Simultaneous global application through a multi-regional clinical trial based on a bridging strategy

There was a growing consensus among pharmaceutical companies that the ICH E5 Guideline needed to be reviewed or a new guideline needed to be developed to enable international clinical trials or simultaneous development to resolve drug lag. Then, the basic concept of the multi-regional clinical trial (MRCT) in which multiple regions participated for the purpose of bridging (utilization of trial data among regions and for application for approval) was presented as the 11th Q&A of the ICH E5 Guideline[10]. Based on this approach, in 2007, the MHLW issued the "Basic principles on Global Clinical Trials" ("Basic Principles"), which outlined the basic principles for planning and conducting MRCT involving Japan for the purpose of application for approval in Japan. The basic principles stated that "A global trial should be designed so that consistency can be obtained between results from the entire population and the Japanese population, and by ensuring consistency of each region, it could be possible to appropriately extrapolate the result of full population to each region." In addition, it introduced 2 approaches to calculate a Japanese sample size considering the possibility to obtain consistent results between the entire population and the Japanese population when designing a MRCT, and the Japanese sample size in subsequent MRCTs was calculated based on the Basic Principles. It was often stated in the review report that PMDA evaluated the efficacy in Japanese patients, in terms of the consistency between the overall study population and the Japanese subpopulation, according to the Basic Principles on Global Clinical Trials and the Basic Principles on Global Clinical Trials (Reference Cases), and this indicated that the evaluation of the data of the Japanese subpopulation and the consistency of the results with the overall population are important in the review for approval. Since Basic Principles was issued, Japanese participation in MRCTs has continued to increase. The proportion of CTNs related to MRCTs was 7.5% in FY 2007, but reached a majority in FY2008, and the number of new drugs approval based on MRCTs has also increased since 2007, and about 40% of new drugs were approved based on MRCTs in FY2018.[49]

ICH Guidelines for Multi-Regional Clinical Trials Assuming Simultaneous Applications for Approval in Multiple Regions

Since around 2000, as the cost of drug development has risen, more and more countries from Asia, Latin America, Eastern Europe, and other regions with lower cost per subject have been participating in MRCTs. As the number of participating countries increases, the number of

[49] Workshop on December 9, 2019 "ICH E17 Guidelines: general principles for the planning and design of global clinical trials" Document of the day (https://www.pmda.go.jp/review-services/symposia/0101.html; final access confirmed May 25, 2022)

subjects in each country (region) becomes smaller accordingly. As a result, it becomes difficult to obtain reliable estimates of the treatment effect in one's own country (region), which is important for assessing the similarity of the results between the country (region) and other regions or in the overall population. If many countries/regions participating in MRCT continue to have a strong interest in the data from their own country/region and try to secure a sample size from their own country/region, the MRCT would collapse. Thus, with the increasing globalization of drug development, there is an increasing need to provide internationally harmonized and general principles for planning and design of MRCT to increase the acceptability of MRCTs in regulatory applications in each region of the world, and it was decided to include MRCT as a topic for ICH. Finally, the ICH E17 Guideline was issued to provide guidelines on MRCTs aiming for simultaneous submission for approval in multiple regions in 2018. The E17 Guideline introduced some tips to scientifically respond to the difficulties we have faced in recent years.

The most important aspect of a confirmatory MRCT is to show the overall treatment effect, but it is also important to evaluate the variability of the treatment effect among regions and intrinsic and extrinsic factors. The E17 approach is to share not only the estimate of the treatment effect on the entire population, but also results of analysis on ethnic factors affecting the treatment effect with each country/region (global first). In other words, when interpreting the results of each country/region, we are not only interested in the results in our own country/region and rely on the estimates obtained from the subpopulation (local first), but also try to use data from the entire population.[50] For example, E17 Training material explained that "if clinically relevant differences among regions are observed, then the MRCT provides a unique opportunity to collect information for additional learning about the factors that may explain these differences (differences of treatment effect among regions)."[51] The E17 Guideline also proposed pooled regions. Pooled regions are pooling countries that share the commonality in intrinsic and extrinsic factors and their distributions in relation to the disease and/or drug under study.

---

[50] Osamu Komiyama, E17 Moves from three regional Clinical Trials to Global Clinical Trials: a Paradigm Shift in New Drug Development and Evaluation is Imminent, Pharmaceutical and Medical Device Regulatory Science 49(5) 295-301 (2018) (https://www.pmrj.jp/publications/02/pmdrs_ topics/topc49-05_ICH-E17.pdf; last access confirmed May 25, 2022)

[51] Training Materials on "General Principles for the Planning and Design of Global Clinical Trials" (https://www.pmda.go.jp/int-activities/int-harmony/ich/0022.html; last access confirmed May 25, 2022)

While the evaluation of the consistency of treatment effects across regions has focused on explaining that the treatment effects across regions are consistent, in pooled regions, it would be to check whether the expected differences have actually occurred because it is expected that there would be differences in the treatment effects across regions based on the known evidence on the factors used in the definition of pooled regions. This is a major difference from the previous evaluation. As for measures to estimate regional effects if the sample size in a region is so small that the estimates of treatment effect will likely be unreliable, the E17 Guideline also introduced a search for options for additional pooling of regions based on commonalities, or borrowing information from other regions or pooled regions using an appropriate statistical model.

Conclusion

The above is a review of the history of drug development, from the independent drug development conducted in each country to today's drug development aiming at simultaneous global development and simultaneous submission by participating in MRCTs. This is a path toward shortening the drug development period and delivering better drugs to patients around the world more quickly by promoting global harmonization and standardization of regulations and standards for drug development, reducing duplication of clinical trials in global and accepting foreign data. On the other hand, it could be considered as a journey from the era in which we had rich data from our own country to the current reality of that data from our own country in clinical data package at submission has been reducing and we have to change our accustomed ways of evaluation and thinking accordingly. We have faced racial and regional differences in accepting foreign data, and have been struggling to evaluate the treatment effects based on a scientific understanding of these issues for many years. This experience is reflected in the ICH E17 Guideline.

The ICH E17 Guideline gives us basic ideas of how to address these difficulties scientifically.

(Back to the main text) (Back to the table of contents)

Column: Argument Based on Data From a Small Number of Cases

International clinical trials that incorporate a very small number of Japanese subjects in drug development are sometimes planned or actually conducted. From a global perspective, it is possible that inclusion of subjects with diverse backgrounds may lead to the discovery of effect modifiers, and that the allocation of the number of cases to many countries and regions may result in the inclusion of a small number of Japanese subjects. Such an approach is **not an E5 approach**, but one based on the E17 concept. For companies, there may be a strategy within the organization, especially through the inclusion of Japanese subjects in early-stage development trials, to create the groundwork for Japanese development personnel to establish a certain status and good relationships within the global development team. However, this is not a scientific reason, but rather a question of how Japanese developers participate and contribute to the global development team and the positioning of the Japanese development organization within the global organization.

What I would like to discuss here are the scientific implications of these cases, where the E5-like approach is the basis but incorporates a small number of Japanese cases, or where the collection of additional data from a small number of Japanese cases does not make a significant contribution to the global drug evaluation. In these cases, it is not uncommon for statistical uncertainty to be underestimated or for discussions to be based on a small number of cases of Japanese data that are scientifically questionable.

Even if the number of cases is small, there may be a vague sense of security in seeing data collected from the Japanese population, such as

> **The Japanese subjects who participated in the global clinical trial were a group with environmental factors (extrinsic ethnic factors) and patient factors (intrinsic ethnic factors) unique to Japan, and the data obtained from this group are unmistakably Japanese data, and the results can be considered to reflect the treatment effect in the Japanese population in general to some extent. The results can be considered to reflect to some extent the effect of the treatment in the Japanese population in general.**

What supports this reassurance is the notion that even if there are unknown environmental or patient factors that are not known to be effect modifiers, the data obtained from Japanese subjects are data that reflect such unknown factors as well.

Even if there are unknown factors characteristic of the Japanese population, information from a Japanese population of several hundred or several thousand cases may be available, as in the case of chronic diseases, but the issue here is the case of a small number of Japanese cases. Even in such cases, unknown factors characteristic of the Japanese population can be easily demonstrated by subgroup analysis of various factors, including candidate effect modifiers in a certain study, as shown in Figure 4, or by comparison of the results of a study conducted in a Japanese population vs. the results of a similar study conducted in a non-Japanese population. In anticipation of such intra and intertrial comparisons, it will be necessary to consider whether the differences to be detected can be detected in the planned trial. For example, in a Phase I study with a Japanese population of a few dozen cases at most, or in a global clinical trial with a small Japanese population, it is not possible to detect differences between the non-Japanese population and the Japanese population.

- ✓ Differences in efficacy endpoints and PK/PD considerations
- ✓ Clinically noteworthy differences in the incidence of adverse events
- ✓ Clinically noteworthy differences in discontinued cases, especially those due to adverse events

The purpose of this study is to determine the probability of detecting the following: What is the probability that such **unacceptable differences can be observed in the number of** Japanese patients in the planned study after considering such differences in advance? What is the probability of observing such a difference with the planned number of Japanese cases? If, after such a study, the differences that should be detected cannot be detected at all, it would be scientifically and ethically unjustifiable to conduct the study, and would be contrary to ICH's objective of avoiding unnecessary duplication of clinical trials in international joint development.[52]

(Back to the main text) (Back to the table of contents)

[52] ICH, Overview of ICH, December 2021 (see p. 4), (https://admin.ich.org/sites/default/files/2021-12/OverviewOfICH_2021_1202.pdf; last access confirmed May 25, 2022) May 25, 2022)

Column: Who is Japanese?

When trying to enroll Japanese subjects in a clinical trial, definitions such as "Japanese people are those who were born in Japan, whose parents are Japanese, and whose maternal and paternal grandparents are Japanese" are used, but since "Japanese" is included in these criteria, the definition are circular and confusing. Considering that grandparents are also determined according to this definition, even the grandparents of the grandparents (second great-grandparents) must be Japanese, and furthermore, in order to show that the second great-grandparents are Japanese, the grandparents of the second great-grandparents... Now, who is Japanese?

There is no simple solution to this question.

Article 2 of the Nationality Law states that a child is a Japanese citizen in the following cases: (i) if the father or mother is a Japanese citizen at the time of birth; (ii) if the father died before the child's birth and was a Japanese citizen at the time of death; or (iii) if born in Japan and both of the parents are unknown or are without nationality. From the perspective of human history, it is widely believed that admixture of the Jomon people, who lived in Japan more than 3,000 years ago, and the Yayoi people came later from the continent resulted in the current Japanese people. The conditions that Japanese people identify a person as Japanese are not limited to nationality or ancestry. A recent sociological study investigated the importance of conditions such as consciousness of being Japanese, place of birth, Japanese language, place of residence, respect for the Japanese legal system, and belief in Buddhism and Shintoism in the identification of Japanese.[53] Who is Japanese? This question may be answered from a variety of perspectives, including anthropology, nationality law, or biology, to define the Japanese for each purpose.

Returning to drug development, the purpose of drug development is to provide effective and safe drugs to those in need of treatment. If the clinical trial data required for regulatory approval consists only of studies conducted in Japan, there is no awareness of the definition of Japanese since the clinical trials are conducted in patients who are likely to use the drug

[53] Akira Igarashi, Research Note, General Image of Japanese Condition: Examining the Ranking of Importance of Conditions by Mokken Scale Analysis, *Theory and Methods* **30(2)** 293-306 (2015) (https://www. jstage.jst.go.jp/article/ojjams/30/2/30_293/_pdf/-char/en; last access confirmed May 25, 2022)

in Japan. In fact, in the Q&A by the Japan Pharmaceutical Manufacturers Association (JPMA) regarding the participation of foreign subjects in clinical trials conducted in Japan, it is stated that the reason why measures have not been taken to avoid the entry of foreign subjects is that "the necessity of Japanese subjects" is not clearly indicated in the regulations, and that clinical trial results conducted in Japan are uniformly accepted as materials for evaluating efficacy and safety in Japanese population as long as they are collected and prepared in compliance with GCP and conform to reliability standards.[54] On the other hand, the background of the target population will be more diverse in the development of a clinical data package that includes MRCTs involving Japan or trials conducted outside of Japan, as clinical trial data required for regulatory approval. The Q&A No. 26 of ICH-E5, for example, mentions the place of birth, generation, diet, and environment, as points to be considered when accepting data from clinical pharmacology studies conducted overseas as Japanese data.[10] The difference between Japanese and non-Japanese, i.e., the definition of Japanese subjects came into focus.

In accepting clinical pharmacology study data conducted overseas as described above, differences in genetic polymorphisms of metabolic enzymes and body size that affect pharmacokinetics are of interest, and there is a need to define Japanese from more biological aspects, i.e., intrinsic ethnic factors. In this case, although it is necessary to consider the influence of differences in extrinsic ethnic factors such as diet and other living environment, a study defining "both grandparents as Japanese" may be sufficient to obtain the desired data. On the other hand, in a MRCT in the late stages of clinical development, extrinsic ethnic factors such as diagnostic methods, medical environment such as standard treatment, and being under Japanese society and customs will be also important in defining Japanese subjects.

These intrinsic and extrinsic ethnic factors can affect treatment effects (modify the effect of a drug) and may be associated with intergroup or interindividual differences in drug efficacy or safety. Individual humans are composed of a variety of factors, and individual regions are also composed of humans with a variety of factors, as well as influenced by extrinsic ethnic factors. The factors that characterize each region may differ from region to region, or they may be common among regions (see figure below). Populations in each

[54] Japan Pharmaceutical Manufacturers Association, Committee on Drug Evaluation, Clinical Evaluation Subcommittee, Clinical Trial 119 Response Team. (2) Entry of foreign subjects. Clinical Trial 119 Questions and Opinions. December 2004. (https://www.jpma.or.jp/information/evaluation/tiken119/02.html; last access confirmed June 1, 2022)

region, including the Japanese, are considered to be characterized by the distribution of these factors, which have commonality and specificity within and among populations, respectively, and the Japanese are one such population. The study of the effects of drugs in different populations then leads to knowledge of the factors that modify the treatment effects. Knowledge of effect modifiers and the degree of their influence will help to estimate the benefit-risk balance based on these factors and ultimately lead to the provision of appropriate medical care tailored to the effect modifiers possessed by individual patients.

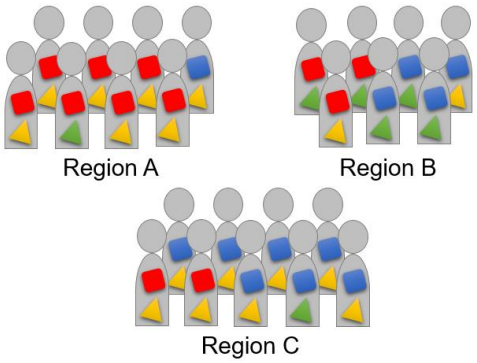


Figure Schematic of regional differences in effect modifiers

There is a difference between the two regions in that Region A has more red for the factor of square, while Region C has more blue. On the other hand, the triangular factor is common to both Regions A and C, with yellow being the most common color in both regions.

(Back to the main text) (Back to the table of contents)

Column: What We Have Seen in the Emergency of the COVID-19 Pandemic

In the emergency situation of the spread of a new type of coronavirus (SARS-CoV-2), we summarized the treatment of Japanese subjects in Japan in emergency situations that have emerged in the development of therapeutic agents and vaccines for the disease.

New coronavirus (SARS-CoV-2) **therapeutic drug**

During the early stages of the spread of the new coronavirus, two drugs, remdesivir and baricitinib, were approved for treatment. Remdesivir was used under the special exception approval system.

What is special exception approval?

The system is based on Article 14-3, Paragraph 1 of the Act on Quality, Efficacy and Safety Assurance of Pharmaceuticals, Medical Devices and Other Products. The system provides for special exception approval for drugs that meet the following requirements: (1) the drug must be used urgently to prevent the spread of disease, (2) there is no other appropriate way to use the drug, and (3) the drug is approved for sale in foreign countries.

Clinical Trial Overview

Generic name	Remdesivir	Baricitinib
Approval Date	May 7, 2020 (special exception approval)	April 23, 2021
Clinical study	International Phase III study (including Japan) (10 countries) Target number of subjects: 400 Preliminary efficacy of 1063 subjects at the time of review No safety information was available as a result of the analysis.	International Phase III study (including Japan) 8 countries, but eventually only one Japanese was in the placebo group Target number of subjects: 1032 (516 in each group) Number of subjects for efficacy analysis: 1,033 (515 in the investigational drug group, 518 in the placebo group) Number of subjects for safety analysis: 1,016 (507 in the investigational drug group and 509 in the placebo group)
	International Phase III study (including Japan) 15 countries, but Part A did not include Japan Target number of subjects: Part A: 400 Part B: 2000 Preliminary figures were available for Part A at time of review	

	Dosing experience conducted from a humanitarian standpoint 9 Japanese subjects treated with the investigational drug in Japan	

In the special exception approval of lemdecivir, the clinical trial at the time of review did not include Japanese subjects.
In the case of baricitinib, which has been approved under the normal review, only one Japanese subject was in the placebo group and no actual drug was administered, but the following observations have been made for the approval of baricitinib.

(Excerpts from the examination report: some changes)
There are no major differences between Japan and overseas in the symptoms of SARS-CoV-2 infection, treatment methods for patients, or risk factors for aggravation, and although the target diseases are different, this drug is effective for rheumatoid arthritis and atopic dermatitis. There is a certain level of usage in Japan as an approved drug, and the dosage and administration of the study was the same as the dosage and administration of the approved drug. Based on the study results, it was judged possible to evaluate the clinical usefulness of this drug to a certain extent in patients with SARS-CoV-2 infection, including Japanese patients.

New coronavirus (SARS-CoV-2) Vaccine

At the time this column was prepared, three drugs developed foreign countries had been approved as vaccines for the new coronavirus.

Concepts on the evaluation of the new coronavirus (SARS-CoV-2) vaccine (at that time) Candidate vaccines developed foreign countries Even in subjects where development has proceeded in foreign countries and large-scale clinical trials to evaluate efficacy and safety have been conducted overseas, the concept of evaluation is the same as for vaccine candidates developed domestically, but in principle, clinical trials should be conducted in Japan to confirm efficacy and safety in the Japanese population. It is advisable to consult with the Pharmaceuticals and Medical Devices Agency as early as possible regarding the necessary domestic clinical trials.

Clinical Trial Overview

Product name	COMIRNATY intramuscular injection	Spikevax Intramuscular Injection
Approval Date	February 14, 2021 (special exception approval)	May 21, 2021 (special exception approval)

Clinical study	Domestic Phase I/II Actual drug group: 120 subjects, Placebo group: 40 subjects Study objectives: safety, tolerability, immunogenicity	Domestic Phase I/II Actual drug group: 150 subjects, Placebo group: 50 subjects Study objectives: Safety, immunogenicity
	Foreign countries Phase I/II/III Study Phase I Part:. Each dose of actual drug or BNT162b1 and each age group: 12 subjects per group Placebo group: 3 subjects per group Objectives of the study: safety, tolerability Phase II/III Part:. Actual drug group: 21,999 subjects; Placebo group: 21,999 subjects Objectives of the study: Efficacy, safety	Foreign countries Phase I Actual drug 25-, 50-, 100-μg group: 35 subjects per group, 250-μg group: 15 subjects per group Objectives of the study: Safety, immunogenicity
		Foreign countries Phase II Actual drug 50- and 100-μg groups: 200 subjects in each group, placebo group: 200 subjects Objectives of the study: Safety, immunogenicity
		Foreign countries Phase III Study Actual drug group: 15,000 subjects, Placebo group: 15,000 subjects Objectives of the study: Efficacy, safety, immunogenicity

Product name	Vaxzevria Intramuscular Injection
Approval Date	May 21, 2021 (special exception approval)
Clinical study	Domestic Phase I/II Actual drug group: 192 subjects, Placebo group: 64 subjects Study objectives: safety, tolerability, immunogenicity
	Foreign countries studies (Combined analysis of Phase I/II [2 studies], Phase II/III and Phase III studies). One dose of actual drug: 1834 subjects, Two doses of actual drug: 11977 subjects Meningococcal vaccine 1 dose group: 1762 subjects Meningococcal vaccine twice or meningococcal vaccine + placebo group: 10200 subjects Objectives of the study: Efficacy, safety, immunogenicity

A similar review policy has been followed in the development of these vaccines.

(Excerpts from the examination report: some changes)
Although at this point in time, no alternative efficacy index for COVID-19 has been identified, and the relationship between efficacy and immunogenicity is not clear, ETIC has decided to evaluate the efficacy of this product in Japanese subjects based on the results of foreign countries validation studies, and to confirm the immunogenicity and safety of this product in Japanese subjects based on the results of clinical studies in Japan. In addition, we decided to evaluate the efficacy and safety of this product in Japanese subjects by confirming its immunogenicity and safety in Japanese subjects based on the results of clinical studies in Japan.

The efficacy and safety of both vaccines have been evaluated in Japanese subjects based on domestic clinical trials, albeit in a small number of subjects.

Future development of therapeutics and vaccines

In the development of therapeutic agents and vaccines for new coronaviruses in Japan, opinions were expressed from various quarters. In particular, there were cautious arguments against domestic approval of vaccines in the early stages because a new type of vaccine called messenger RNA had been developed, but later opinions were expressed from various quarters calling for a review of the delay in vaccine approval compared to other countries. For example, the urgent approval should be based on safety and efficacy data from other countries such as Europe and the US,[55] and it is difficult to make a statistical argument about efficacy based on data from about 160 Japanese, and it is impossible to detect serious adverse reactions as frequent as 1 in 100,000 inoculations.[56]

In parallel with these discussions, the Japanese government took the lead in continuing the study and promulgated the "Act for Partial Revision of the Act on Quality, Efficacy and Safety Assurance of Drugs and Medical Devices (Act No. 47 of 2022)" to prepare for a new pandemic of infectious diseases, taking into account that the vaccination against new coronaviruses was delayed compared to Europe and the US. (Law No. 47 of 2022) and the related notice[57] was

---

[55] Hiroshi Mikitani, New Economic Federation Urgent Recommendations for Vaccine Measures to Overcome the Corona Problem (https://jane.or.jp/proposal/comments/14036.html; last access confirmed May 25, 2022)

[56] Nikkei Business Editorial Board Member Nikkei Biotech Editorial Board Member Muneaki Hashimoto (former Ministry of Health, Labor and Welfare official)
Is the Corona Vaccine "Clinical Trial in Japan" Scientifically Correct?

[57] Approval Examination Approach under the Urgent Approval System, Pharmaceutical and Pharmaceutical Affairs Ministry Hearing No. 0520-1, May 20, 2022.
(https://www.mhlw.go.jp/content/11120000/000940766.pdf; last accessed May 25, 2022)

issued at the same time. The Act expands the scope of the emergency approval system to include not only vaccines and therapeutic drugs, but also pharmaceuticals in general, medical devices, and regenerative medical products. The Act assumes that safety will be confirmed to the same level as that of regular regulatory approval, but also that efficacy will be confirmed to the same level as that of regular approval under the condition that additional information is submitted after approval. However, the efficacy of the product is not yet confirmed at the usual level, with conditions such as the submission of additional information after approval. We believe that the content of the proposal is unprecedented.

Summary

The development of vaccines and therapeutics against SARS-CoV-2 was an urgent societal need that would have been unthinkable under normal circumstances. Even under such circumstances, it is true that the applicant, the regulatory authorities, and even the general public were convinced that obtaining data from Japanese nationals was important in the development of these vaccines and therapeutics. Although the framework of special exception approval was applied, and as a result the situation was resolved, the idea that obtaining data from Japanese subjects is important for the evaluation of drugs for the Japanese population remained the Golden Standard, and even in emergency situations it was basically unwavering. For a drug such as a vaccine that targets a large number of unaffected people before they become ill (the entire population, since anyone could have been infected with SARS-CoV-2), it seemed extremely important to be able to predict what would happen in the Japanese population. There may have been some concern or vague fear that an unknown mechanism might cause adverse reactions specific to the Japanese population that had not been anticipated.

However, if we could think in terms of effect modifiers (risk factors for adverse reactions), as this report discusses, the development plan and review might have been different. What was observed in human beings around the world, and beyond the interpretation of trials conducted in specific regions, what were the benefits and risks to people of different backgrounds? Can we characterize the Japanese population with the effect modifiers focused on this? As a result, what are the expected benefits and risks in the Japanese population? From this perspective, to what extent can the results of existing clinical trials conducted in foreign countries be used to make claims about efficacy and safety, and what should be supplemented? When considered from such perspectives, we should look back while referring to the notice issued this time, and organize the Lessons Learned from the emergency. Considerations in terms of efficacy and safety are as follows: the extent to which

the results of existing clinical trials conducted in foreign countries can be claimed; what should be supplemented; how can we fill missing information (in other words, Missing Pieces of Knowledge) with newly collecting Japanese data; and if the development plan and regulatory review based on these considerations. In addition, in the future, the notice issued this time may have some points in common with the presentation of results incorporating the concept of effect modifiers discussed in this paper, and we would like to conclude this section with the hope that such a concept will spread even outside of emergencies.

(Back to the main text) (Back to the table of contents)