CTD Structure Using the Results of Multi-Regional Clinical Trials Based on the Principles in ICH E17

Japan Pharmaceutical Manufacturers Association
Data Science Subcommittee, Drug Evaluation Committee
Task Force 2-2 for FY 2022
E17 Subteam

July 25, 2023

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviations	Unabbreviated term
ARR	attributable risk ratio
bid	bis in die
BMI	body mass index
CART	classification and regression tree
CI	confidence interval
CTD	Common Technical Document
DOR	duration of response
EDC	electronic data capture
EMA	European Medicines Agency
EU	European Union
FAS	full analysis set
FDA	Food and Drug Administration
GCP	good clinical practice
HDI	human development index
HR	hazard ratio
ICH	International Council for Harmonization of Technical Requirements for Pharmaceuticals for human use
ITT	intention-to-treat
KM	Kaplan-Meier
LS	least squares
MedDRA	Medical Dictionary for Regulatory Activities
MRCT	Multi-regional Clinical Trial
NMPA	National Medical Products Administration
ORR	overall response rate
OS	overall survival
PFS	progression-free survival
PhRMA	Pharmaceutical Research and Manufacturers of America
PK	pharmacokinetics
PMDA	Pharmaceuticals and Medical Devices Agency
PT	preferred term
Q-Q Plot, Q-Q plot	quantile-quantile plot
RR	relative risk
SE	standard error

1 Introduction

After ICH published E5 [1)] in 1998 and the "Guideline for Ethnic Factors in the Acceptability of Foreign Clinical Data" (Iyakushin No. 672) was implemented in Japan in the same year, many new drugs were approved for marketing based on a sequential bridging strategy (Uyama et al. 2005). Subsequently, ICH published E5 Q&A No. 11 in 2006. This Q&A provides guidelines on the design, conduct, and reporting of multi-regional clinical trials (MRCT) within the framework of E5 and perform a certain function as a preliminary step to the MRCT that are described in ICH E17. Since the “Basic Principles on Global Clinical Trials” (PFSB/ELD Notification No. 0928010) was issued in 2007, the MRCT and global development in which Japan participates have been accelerated. The mid-2000s coincided with the time when the principle for international mutual use of clinical study data, which was built up through E6 (GCP) agreed by ICH in 1996, began to bear fruit, and the IT environment to support electronic data collection represented by EDC and global development began to be established. Thus, MRCT involving countries other than Japan, the US, and Europe rapidly increased worldwide. Due to this increase, many regulatory authorities including those in Japan have had to consider the acceptance of data such as race, ethnic factors, or medical environment from other countries. In view of these issues arising from regulatory authorities and the pharmaceutical industry's accumulated experience in drug development, ICH discussed the E17 guideline for the aim of increasing the acceptability of MRCT for marketing authorization applications across countries/regions, and agreement was reached in 2017. The ICH E17 Guideline, "General Principles for Planning and Design of Multi-regional Clinical Trials" (PSEHB/PED Notification No. 0612-1) was issued in Japan, and the guideline was implemented.

ICH E17 focuses on the planning and design of MRCTs and provides considerations for ethnic factors that are important to medical products and for designing MRCTs appropriately based on regional differences in these factors. The plan and design of a clinical trial are to be determined after careful consideration based on prior knowledge and the perspective of assessing the study results. Therefore, it is important to design studies based on a “global-first” concept that evaluates the efficacy and safety of a medical product in the country/region of application after investigating the influences of ethnic factors on the efficacy and safety of the investigational drug from various perspectives and systematically based on the results of all MRCTs planned and conducted in accordance with ICH E17. In addition, enhanced evaluation based on overall MRCT results will enable the results to be used globally, reduce the time to approval by reducing the number of discussions that have been conducted individually in a particular country/region, and accelerate worldwide access to medicine.

Although many MRCTs were conducted after the implementation of ICH E17 and many new drugs have been approved for marketing based on the results, few Common Technical Documents (CTD) are prepared based on ICH E17. As in the past, most CTD are written from a “local-first” concept based on ICH E5, which focuses on the results in the application country/region first, and then evaluates the similarity with other regions or the whole study. We believe that misunderstanding of ICH E17 and lack of understanding of the differences between ICH E5 and ICH E17 have also influenced this approach.

---

1): ICH E5 was approved by ICH only 1 month after reaching Step 4 agreement at ICH in February 1998 with minor modifications and has since undergone 1 revision (R1; Revision 1). In this document, E5(R1) is abbreviated as E5.

In this document, we will explain the misunderstanding of ICH E17 and the differences in basic concept between ICH E5 and ICH E17, describe ICH E17in Section 2, describe the framework of the 3-layer approach in Section 3, and propose a strategy for the structure and contents of CTD based on the principles of ICH E17 in Section 4. This document focuses on the evaluation method of MRCT results and how to summarize them in a CTD, intends to deepen the understanding of multifaceted and systematic evaluation methods based on the principles of ICH E17, and proposes a strategy for evaluation, interpretation, and discussion of results. We hope that this document will contribute to the interpretation of MRCT results and the preparation of CTD based on the principles in ICH E17.

1.1 Misunderstanding of ICH E17

Frequently asked questions about ICH E17 include:

- Can ICH E17 only be applied when a pooling strategy is used?
- When considering ICH E17, it is inappropriate to set the number of participants in accordance with Methods 1 and 2 of “Basic Principles on Global Clinical Trials” (PFSB/ELD Notification No. 0928010) published in 2007 within the framework of ICH E5 Q&A No. 11?

The answer to both questions is “No.”

ICH E17 is a guideline that provides general principles for the planning and design of MRCTs with the intent to increase the acceptability of MRCTs for registration globally. Section ICH E17 presents general principles for a wide variety of MRCT study designs, including, but not limited, to a pooling strategy. "Consider ethnic factors that are important to the medical product and design MRCTs appropriately based on regional differences in these factors" refers to the application of E17.

However, it is often probably misunderstood that the application of ICH E17 is only the application of the pooling strategy because when ICH E17 was formulated, a strong impression was made about the newly-defined terms of “pooled region” and “pooled subpopulation.” In addition, expectations for the pooling strategy have been raised for the purpose of reducing the number of participants, but it is difficult to predefine the pooling strategy and obtain agreement with the regulatory authorities at the time of MRCT planning, and consequently the misconception has spread that "ICH E17 is not applicable unless the regulatory authorities agree with the pooling strategy at the time of planning the MRCT." Furthermore, the Japanese sample size was determined in accordance with Method 1 or 2 described in Question No. 6 of the 2007 notification (PFSB/ELD Notification No. 0928010) without considering the newly-defined pooling strategy in ICH E17, which often leads to the misconception that ICH E17 is not applicable.

Due to the misunderstanding that ICH E17 cannot be applied, applicants may be stuck in conventional thought based on ICH E5 to focus on the results in the country/region of application (“local first”) and then evaluate the similarities in other regions or in the entire study. This may be the essence of the failure to correctly understand and utilize ICH E17.

1.2 ICH E5 vs. ICH E17

ICH E5 Q&A No. 11, the previous guidance of ICH E17 on the design, conduct, and reporting of multi-regional trials, relied on a local-first concept in which the bridging strategy focused first on the application country/region results, then on the evaluation of similarities across other regions or studies as a whole. It is necessary to clarify once again that ICH E17 is not a revision of ICH E5 or ICH E5 Q&A No. 11 but is a separate guideline, and then deepen the understanding of the multifaceted and systematic evaluation methods based on the principles of ICH E17.

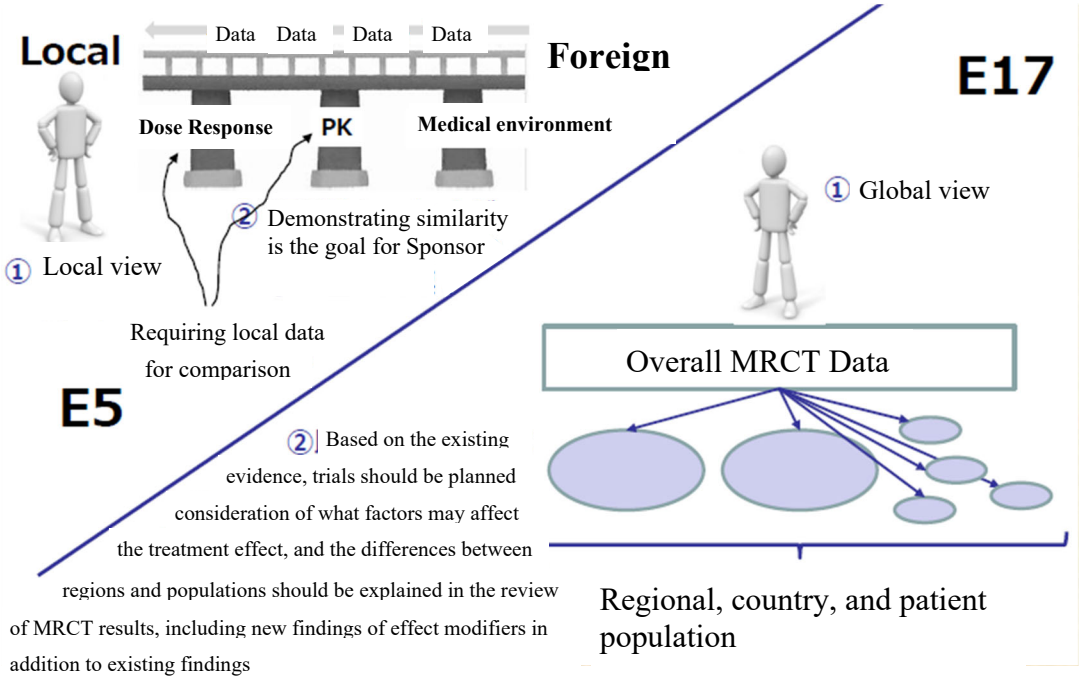
ICH E5

ICH E5 organized the factors that affect the treatment effect as extrinsic and intrinsic ethnic factors and reorganized the points to note for accepting foreign clinical study data by introducing the concept of bridging. The bridging strategy introduced by E5 is essentially first on the application country/region (local). For example, if Japan is local and a clinical data package has already been approved in a foreign country, a bridging study with the same design as the foreign clinical study will be conducted in Japan, and the results will be compared. Then, if the efficacy and safety, dose response, and pharmacokinetics (PK) are similar and the medical environment is also similar to that in the foreign countries, the bridge pier will be completed as shown in Figure 1-1, and the foreign clinical data package will be extrapolated to the Japanese clinical data package. In other words, the main strategy is to show the similarity between Japan and overseas clinical studies in order to utilize the results of overseas confirmatory studies for drugs that have already been approved overseas but are not yet approved in Japan.

ICH E17

ICH E17 is a guideline describing the general principles of MRCT planning and design, which is a strategy based on the entire MRCT study results or the entire clinical data package; in other words, a globally focused strategy. The central principle of ICH E17 is to investigate important ethnic factors that may affect the treatment effect of the medical product, and to plan MRCT based on differences in these factors among regions, and also to investigate newly-found differences in MRCT results among regions and populations, including ethnic factors (effect modifiers) that may affect the treatment effect, in addition to existing findings. Please note that the consistency evaluation described in ICH E17 includes an attempt to explain observed differences between regions or populations in terms of effect modifiers, and the goal is not to show similarities between regions or populations.

Figure 1-1 ICH E5 vs. ICH E17



Source: 7th ICH E17 Workshop (2022)

In evaluations based on the principles of ICH E17, it is important to consider ethnic factors (effect modifiers) that affect the treatment effect. Efficacy modifiers and the details of the evaluation, mainly of effect modifiers, are explained in Section 3.

2 ICH E17

As described above, ICH E17 “General Principles for Planning and Design of Multi-regional Clinical Trials” has been developed to define the general principles of planning and design of MRCT to increase the acceptability of MRCT in the approval application in various regions of the world. It describes the selection of participants, dosage, endpoints, control drugs, and statistical analysis plan that are issues in considering the design and protocol of MRCT.

Training materials have been prepared to promote understanding of the important points of this guideline. Module 1-7 contains the training material, including an overview in Module 1 and detailed descriptions focusing on the respective sections of ICH E17 in Module 2-7 (Table 2-1).

Table 2-1 Overview of training material module 1-7

Module	Title	Outline
Module 1	Basic principles and overview of training modules	• Purpose and overview of training material
Module 2	Considerations of regional variability when recruiting diverse populations in global drug development	• The importance of prior consideration of ethnic factors affecting treatment response in the design of MRCT • Methods to identify intrinsic or extrinsic factors affecting treatment effects and its countermeasures
Module 3	Selection of doses for use in confirmatory MRCTs	• Considerations for dose selection • Introducing cases that different dosing regimens were selected in confirmatory studies (MRCT)
Module 4	Overall sample size and allocation to regions	• Statistical considerations for setting the sample size target • Considerations when considering the sample size in each region • Concept of pooled regions and subpopulations • Introduction of 5 methods of sample size allocation addressed in the guidelines (advantages and disadvantages of each method)
Module 5	Pooling strategies	• Significance of pooling • Definition of pooled region and pooled subpopulations • Methods for pooling regions and subpopulations • Applicability of pooling strategy
Module 6	Evaluation of consistency	• Definition of consistency • Significance of evaluation of consistency • Procedure and considerations for evaluation of consistency • Consider evaluation of consistency by region • Case study: PLATO study
Module 7	Selection of comparators	• Procedure for selection of comparators • Considerations when selecting drugs not approved in a specific region

Source: ICH E7 Training Material Module 1-7

Module 6 links the “Planning and Design” included in the title of ICH E17, which is the basic scope of ICH E17, to the evaluation of the MRCT results on which this document focuses (interpretation and evaluation of results, including consistency across regions and

subpopulations). Since this document intends to focus on how to evaluate the results of MRCT, ICH E17 Training Material Module 6 is described first.

2.1 Training Material Module 6

In Module 6, the basic concepts for evaluating the study results of MRCT are explained as shown in Figure 2-1. The statement at the bottom of this figure **"These eventualities should be carefully considered at the planning stage"** is emphasized by the ICH E17 Implementation Working Group that created this training material, which conveys that the "planning and design" of MRCT should be conducted in consideration of the "interpretation and evaluation of results."

Consistency is defined as "absence of clinically relevant differences between treatment effects in different regions or subpopulations of an MRCT." The goal of the assessment of consistency is not to explain the consistency of results among regions or subpopulations. It also means to evaluate the presence or absence of differences in results among regions or subpopulations from various perspectives and systematically, and if differences are observed, examine whether an explanation can be provided by effect modifiers (see Section 2.2). In other words, it is important to evaluate whether the results of the overall population are applicable to all regions with effect modifiers as the main axis.

Module 6 provides a structured step that carefully considers intrinsic and extrinsic factors and the five perspectives to evaluate consistency (Figure 2-1). A structured step is a stepwise process of exploring effect modifiers. It has been shown that when there are known factors, they should be evaluated after analyses are specified at the time of planning the trial. When unexpected differences between regions or among subpopulations are observed, the factors need to be explored, such as by subgroup analyses. Differences between regions or among subpopulations that cannot be explained by exploration may require further post-hoc analyses, and additional studies may be needed in some cases for exploration. In addition, the five perspectives have been shown to enhance the credibility of MRCT results: biological plausibility, internal consistency, external consistency, statistical uncertainty, and clinical relevance (Figure 2-2).

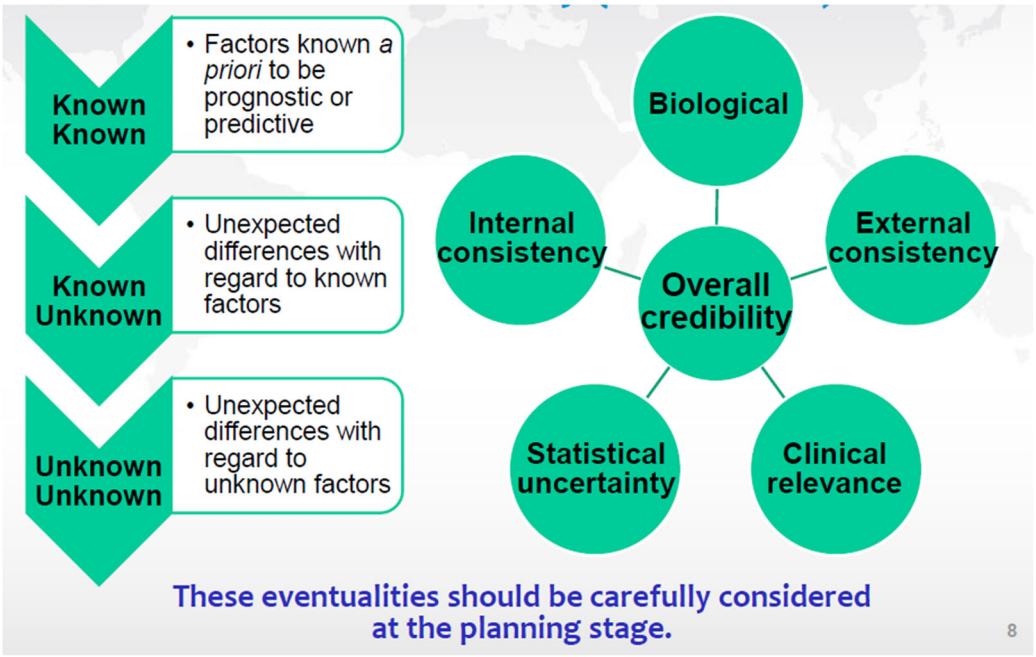
These five perspectives are explained as follows:

Biological Plausibility	Is a biologically compelling explanation possible?
Internal Consistency	Are there multiple findings in the same study that support each other, e.g., are there findings in multiple biologically or medically relevant endpoints that support the same conclusion?
External Consistency	Are the findings from one study aligned with external information, such as other study results, that support the same conclusion?
Statistical Uncertainty	How credible is the finding? In other words, is there any bias in the estimation that cannot be ignored, and how precise is the estimation?
Clinical Relevance	Is it a significant finding that provides the basis for clinical judgment or treatment policy decisions? Are the observed differences among any populations clinically significant?

Source: Drug Evaluation From Now On--- What's the value of local data in the global clinical data package? - (Japan Pharmaceutical Manufacturers Association 2022)

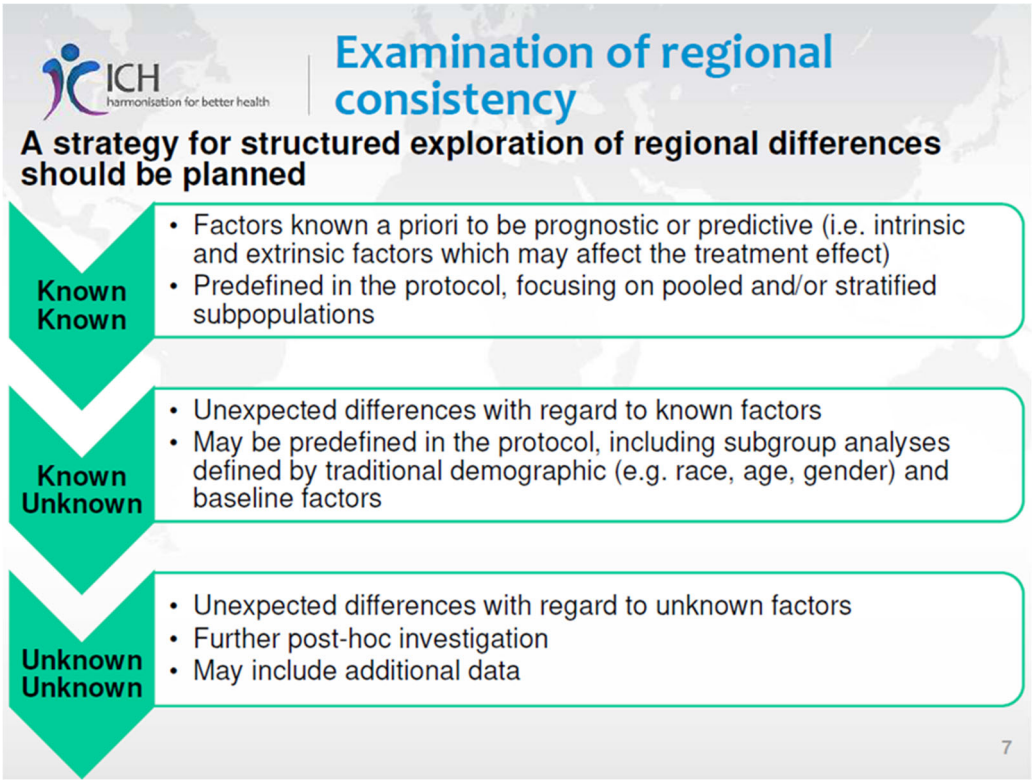
The details are described in Section 3, together with the 3-layer approach, which is a framework for systematic evaluation of MRCT results.

Figure 2-1 Multi-faceted evaluation to increase the credibility of MRCT results



Source: ICH E7 Training Material Module 6, page 8

Figure 2-2 Structured steps in exploring effect modifiers



Source: ICH E7 Training Material Module 6, page 7

2.2 Effect Modifiers

This section describes effect modifiers and their approach.

A biological phenomenon in which the treatment effect (efficacy or safety) of a drug differs is called an effect modification, and factors causing effect modification, that is, factors affecting the treatment effect, are called effect modifiers ["Drug Evaluation Fron Now On ---What's the value of local data in the global clinical data package? -" (Japan Pharmaceutical Manufacturers Association 2022)]. Factors that can be effect modifiers include patient factors (intrinsic factors) that reflect individual patient characteristics and environmental factors (extrinsic factors) that reflect living and medical environment characteristics. Prognostic factors of the disease and predictors of treatment can also be effect modifiers. Also, for the same drug safety and efficacy, there may be the same or different effect modifiers.

Once an effect modifier has been identified, the treatment effect can be interpreted while considering the effect of effect modifiers when interpreting clinical study results. For example, the treatment effect in each subgroup will be evaluated based on the results of subgroup analyses performed on effect modifiers. If multiple effect modifiers are present and there is no relationship between them, separate subgroup analyses can be evaluated. In addition, another approach is to construct subgroups combining multiple factors so that the results of subgroup analysis can explain the combined effects of a combination of effect modifiers.

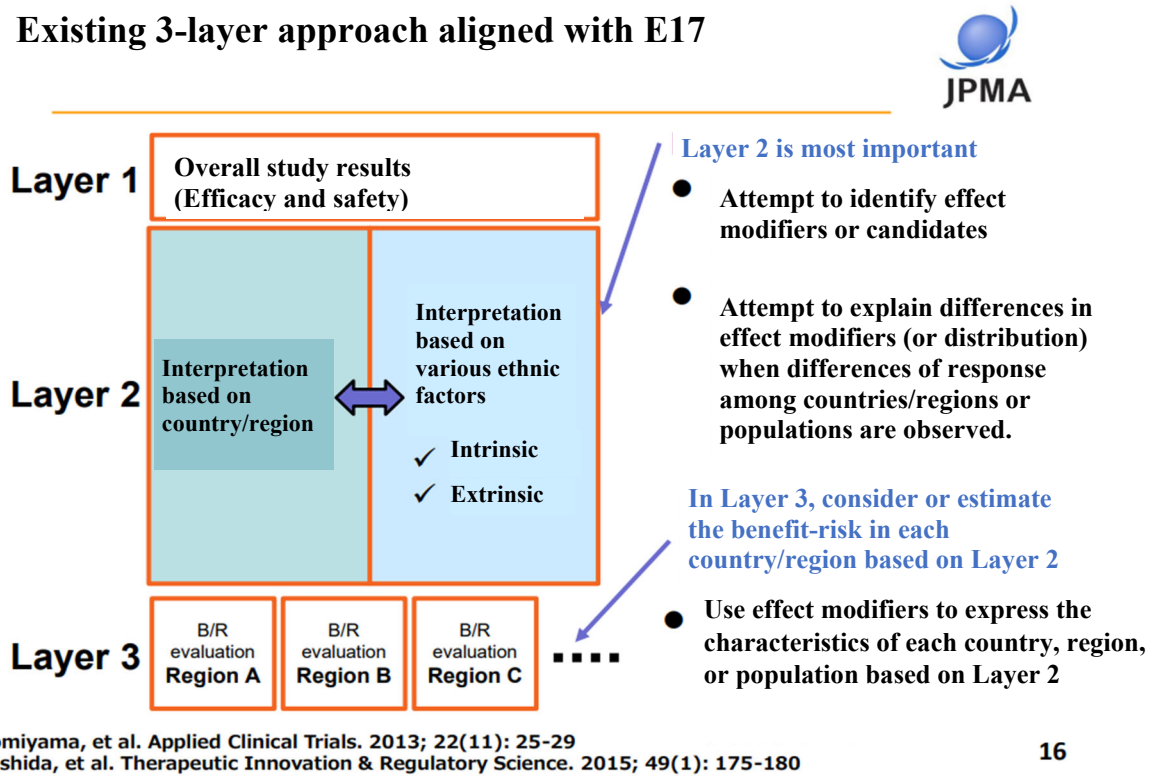
3 3-layer Approach

3.1 Description

This section outlines a 3-layer approach (Komiyama et al. 2013) proposed by Komiyama et al. The 3-layer approach is a framework to evaluate the results of MRCT and is a scientific approach to share and evaluate the results of MRCT as a whole in the country/region of application. Overall MRCT results refer not only to a statistically significant primary analysis of the primary endpoint that achieved the study objectives, but also to the estimates and an analysis of the ethnic factors that influenced the estimates. This is an approach to interpreting the results in country/region of application in the context of the overall results without a focus on the data in country/region of application.

The 3-layer approach (Figure 3-1) is a framework for evaluating the results of multi-regional clinical trials consistent with the principles of ICH E17. Interpretation of MRCT results are divided into three layers, and Layer 1 and Layer 2 are evaluated globally. In other words, based on the overall study results, the efficacy and safety of the medical product are evaluated in Layer 1, and the presence or absence of factors affecting the results, effect modifiers, or candidates are evaluated in Layer 2. If different responses among countries, regions, or populations were observed, it is investigated whether this can be explained by differences in effect modifiers or their distribution. Layer 3 evaluates the benefit/risk in the application country/region based on the results of Layer 1 and 2 using effect modifiers to represent the characteristics of the country, region, and population in which the treatment effect is intended to be estimated. The details of each layer are described in the following sections.

Figure 3-1 Outline of the 3-layer approach



Source: 7th ICH E17 Workshop (2022)

3.2 Layer 1: Evaluation in the Overall Population

Regardless of the country in which the study is taking part, the primary objective of the study is to estimate the treatment effect based on data from all participants, except in special circumstances. In the 3-layer approach, the purpose of Layer 1 is to evaluate and understand the overall results of the MRCT as a single clinical study. Clinical trials are designed to evaluate a primary objective and a study design (target population of study to be reflected in the inclusion criteria, composition setting of the treatment arm, primary endpoint and its evaluation method [including statistical hypothesis structure], number of participants) is determined to achieve its objective. At the stage of evaluation of study results, even in a MRCT, without focusing on a specific country or region, it is first examined whether the primary objective has been achieved and what the results were related to other efficacy endpoints and safety. This assessment provides the basis for the subsequent strata (Layer 2 and Layer 3) and provides a starting point for further evaluation. This idea is also aligned with the principles underlying ICH E17 (see ICH E17 “Basic Principles” in Section 1.4, “General Recommendations in the Planning and Design of MRCTS” in Section 2).

The five perspectives in ICH E17 Training Material Module 6 shown in Section 2.1 are important perspectives to enhance the credibility of the results of not only MRCT but also the overall study, and should be considered at Layer 1. Five perspectives to be considered in Layer 1 are described below, along with some examples of efficacy.

Biological plausibility

Examine whether the results of the overall study can be reasonably explained from multiple and multifaceted viewpoints clinically, pharmacologically, or based on the mechanism of action. Considerations of biological plausibility should also be linked to considerations of internal and external consistency. As an example, the case below is discussed from a clinical and mechanistic point of view.

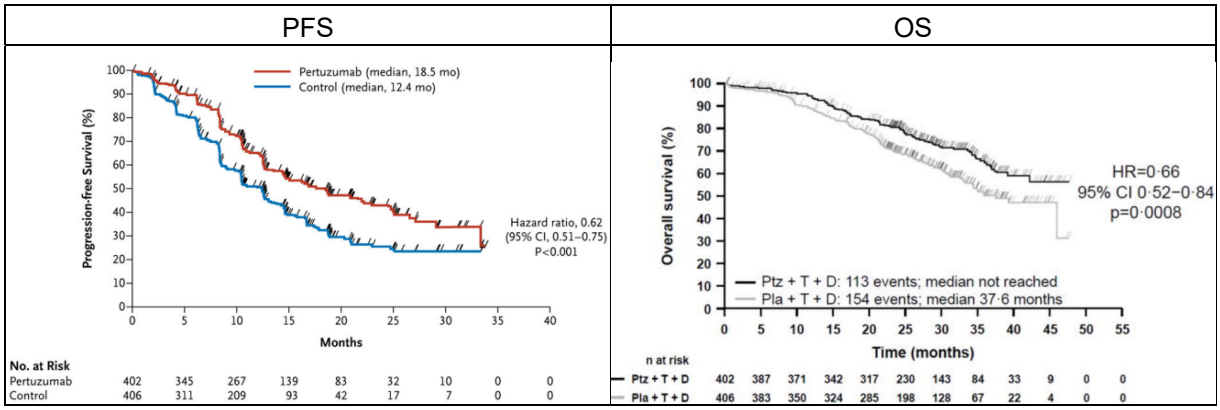
Case example: Fesoterodine fumarate

Fesoterodine fumarate, a muscarinic receptor antagonist, binds to the muscarinic receptor present in the nerve terminals of bladder smooth muscle and vesical sensory nerve to suppress the increase in bladder contraction and micturition reflex. In the multi-regional Phase II study in patients with OAB, treatment with fesoterodine fumarate resulted in a statistically significant decrease (improvement) in the mean frequency of urge urinary incontinence per 24 hours (primary endpoint) compared to placebo. Similarly, the mean number of micturitions and mean number of urgency episodes decreased (improved) compared with placebo. Similar results were obtained in non-Japanese Phase III studies. The results show biologically plausibility with consistent results from a perspective of the clinical viewpoint and the action of mechanism.

Internal consistency

Consider whether multiple results within the same study are supportive of each other, e.g., results of multiple biologically or medically relevant endpoints that support the same conclusion. Evaluating the consistency of results between medically relevant primary and secondary endpoints is one of the examples to show internal consistency. For example, in the case of pertuzumab shown in Figure 3-2, the consistency of results between the medically relevant primary and secondary endpoints (progression-free survival [PFS] and overall survival [OS], respectively) were assessed to see internal consistency.

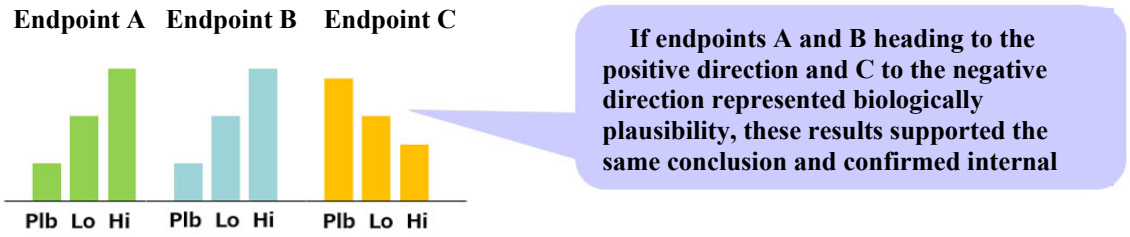
Figure 3-2 Layer 1: Case for internal consistency of efficacy (pertuzumab)



Source: Baselga et al. 2012, Swain et al. 2013

It is also useful to evaluate the summary statistics by treatment group for multiple endpoints as in Figure 3-3.

Figure 3-3 Layer 1: Case for evaluation of internal consistency (hypothetical cases)



Source: 7th ICH E17 Workshop (2022)

External consistency

Evaluate whether the results from one study and external information, such as other study results, support the same conclusions. If there are multiple studies that evaluated the same drug in a similar patient population, the consistency between studies should be evaluated to confirm the robustness of the evidence obtained. If there are no studies that can evaluate the consistency of the same drug, one option is to evaluate the consistency with the results of studies of drugs of the same class and indication.

For safety, comparisons of data obtained when the same drug is studied for different indications may also be useful in an assessment of consistency.

The case example below shows two studies that evaluated the same drug in a similar patient population.

Case example: Brolucizumab (genetical recombination)

Figure 3-4 shows that 2 confirmatory studies (KESTREL and KITE) evaluating the efficacy and safety of brolucizumab in patients with diabetic macular edema (DME) in the same publication (Brown et al. 2022). These 2 aflibercept controlled, noninferiority studies evaluated the consistency of efficacy and safety across studies between the brolucizumab 6-mg group and the aflibercept group. Furthermore, the results were consistent with those of the HAWK and HARRIER studies that evaluated the efficacy and safety of brolucizumab in patients with subfoveal choroidal neovascularization secondary to age-related macular degeneration (nAMD).

Figure 3-4 Layer 1: Case examples for evaluation of external consistency of efficacy (brolucizumab)

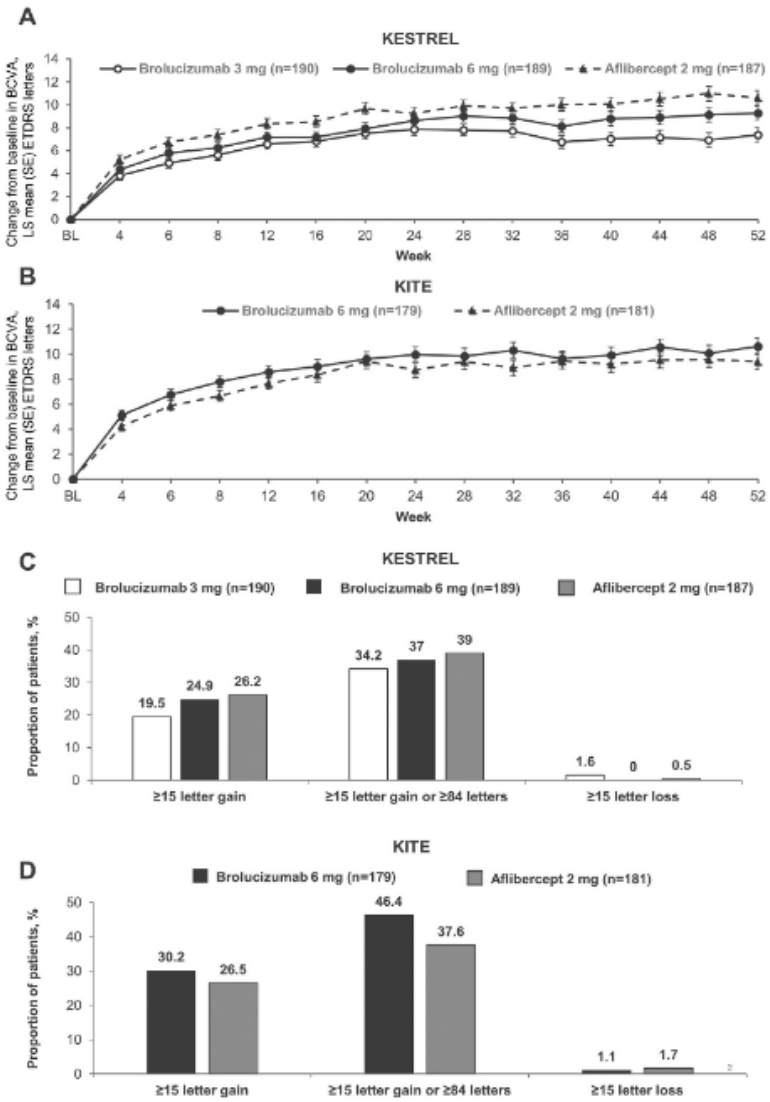


FIGURE 2. Mean change in BCVA from baseline to Week 52 in A) KESTREL and B) KITE. Proportion of subjects with 15-letter gain/loss in BCVA in C) KESTREL and D) KITE at Week 52. Full analysis set, last observation carried forward. BCVA = best corrected visual acuity, BL = baseline, ETDRS = Early Treatment Diabetic Retinopathy Study, LS = least squares, SE = standard error.

Source: Brown et al. 2022

Statistical uncertainty

An assessment of uncertainty is an important and essential component in evaluating scientific evidence.

In Layer 1, the study design (number of participants, blinded or unblinded) and study integrity (status of protocol deviation) were comprehensively evaluated, and the reliability of the results obtained from all participants, if there are unignorable biases in the estimates, and the accuracy of estimation are examined.

Clinical significance

Consider whether it is important findings that provide the basis for clinical decisions or treatment decisions.

In Layer 1, considering whether the results of the overall population, the clinical significance of the results obtained, their generalizability to routine clinical practice, the strength of evidence related to each key benefit and risk, and the results of analyses of limitations and uncertainties are important findings that provide the basis for clinical decisions and treatment decisions.

3.3 Layer 2: Consideration of Effect Modifiers From Various Perspectives and Evaluation of Consistency Among Regions

Layer 2 is the key layer underlying Layer 3, which is the benefit-risk assessment in each country/region of application. In Layer 2, data from all participants in the MRCT will be used to evaluate factors affecting treatment effect, unless otherwise specified. As shown in Figure 3-1, factors affecting the treatment effect will be explored from the perspectives of country/region and various ethnic factors (known effect modifiers or candidates). Forest plots showing subgroup analyses by each factor and their results are the starting point of this exploration. After that, the exploration will be carried out from various viewpoints as explained below. In addition to considering the five perspectives in ICH E17 Training Material Module 6 shown in Section 2.1, when differences in treatment effects among countries or regions are suggested, it will be considered whether the differences can be explained by differences in effect modifiers or their distribution.

On the other hand, attention should be paid when no difference is observed between countries/regions and no candidate effect modifier is suggested even after multifaceted data analysis. Proof that there is no difference in the treatment effect in any patient population (that the whole MRCT result can be extrapolated to any country/region) is so-called “devil’s proof,” or “proof of the negative fact,” and it is necessary to conduct a more multifaceted and elaborate analysis and send careful messaging keeping in mind the adage "Absence of evidence is not evidence of absence."

The following sections describe approaches including the consideration of effect modifiers from various perspectives (Section 3.3.1.1, Section 3.3.2.1) and the evaluation of consistency in

regions (Section 3.3.1.2, Section 3.3.2.2) to evaluate the presence/absence of effect modifiers or candidates by the efficacy and safety evaluation.

The five perspectives shown in ICH E17 Training Material Module 6, Section 2.1 are important perspectives to enhance the credibility of MRCT results and important perspectives to evaluate Layer 2. The points to consider when examining effect modifiers in Layer 2 are explained below.

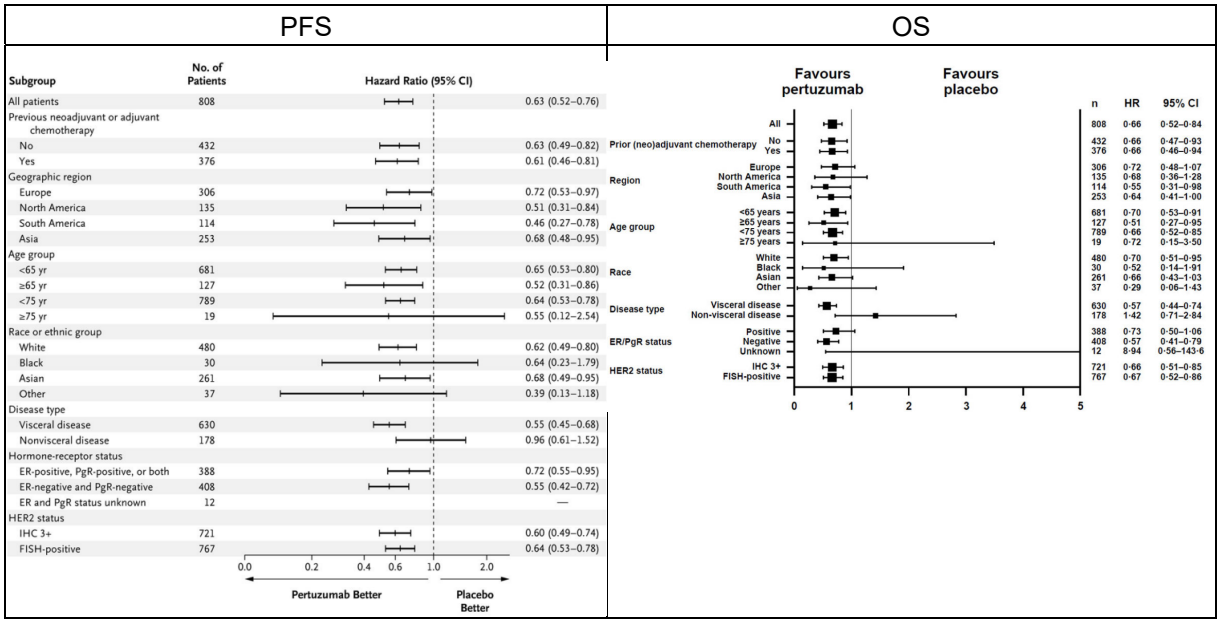
Biological plausibility

Consider whether differences in results observed among regions or populations and effects of candidate effect modifiers on treatment effects can be reasonably explained from multiple and multifaceted viewpoints clinically, pharmacologically, or based on the mechanism of action. Consideration of biological plausibility is also related to assessment of internal and external consistency.

Internal consistency

Consider whether differences in results observed among regions or populations and effects of potential effect modifiers on treatment effects are consistent across multiple biologically or medically relevant endpoints, or across prespecified subgroups (regions or subpopulations). For example, it is useful to evaluate whether the effect of candidate efficacy modifiers on the therapeutic effect of an anticancer drug is consistent between PFS and OS by creating forest plots as shown in Figure 3-5.

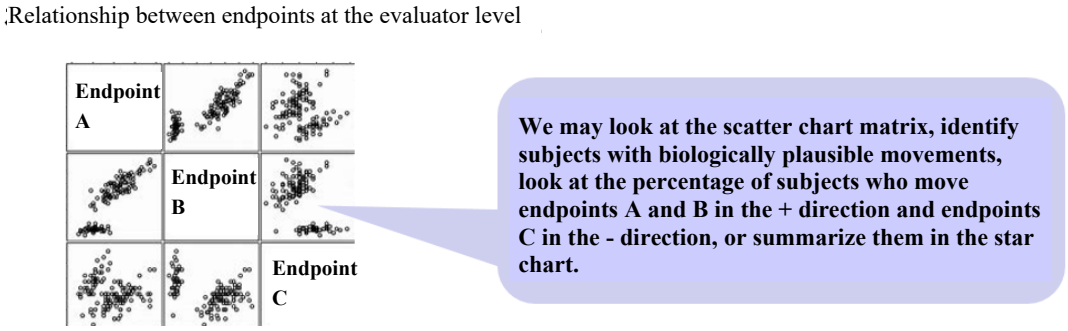
Figure 3-5 Layer 2: Internal consistency of efficacy case (pertuzumab)



Source: Baselga et al. 2012, Swain et al. 2013

As shown in Figure 3-6, it is also useful to plot the results of medically related primary and secondary endpoints individually and evaluate the consistency of the effect of a candidate effect modifier on the treatment effect. When there are multiple endpoints, preparing a scatter plot with all the binary combinations placed next to each other may identify which participants are consistent in terms of the clinical and mechanistic aspects of biological plausibility and which are not. Creating scatter plots by color-coding the scatter matrix per predefined subgroups may identify factors that are common among participants who are not biologically plausible, i.e., potential effect modifiers.

Figure 3-6 Layer 2: Case example for evaluating internal consistency of efficacy (hypothetical cases)



Source: 7th ICH E17 Workshop (2022)

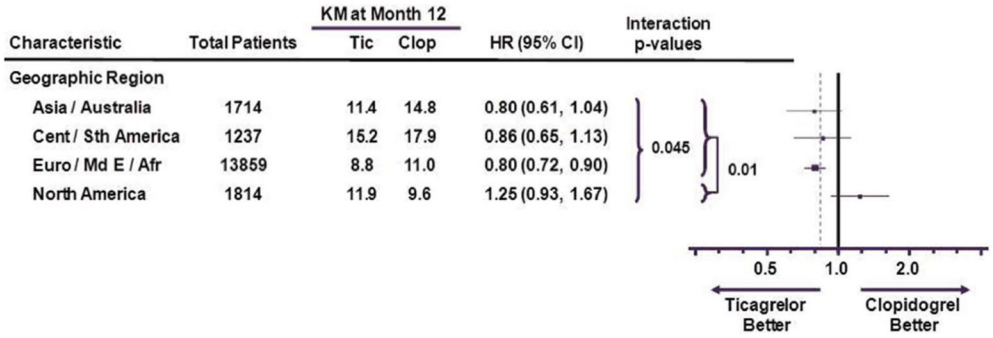
The PLATO study, which is described in ICH E17 Training Material Module 6 as an example of internal consistency across regions, is also described as an example of assessing the impact of candidate effect modifiers on treatment effects across subgroups (regions or subpopulations). Details of the PLATO study are provided in Section 5.1.

Case example: Ticagrelor

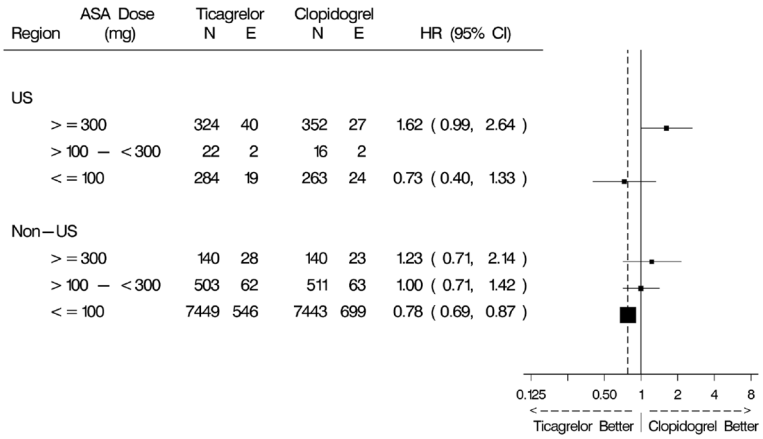
In the PLATO study, the results of subgroup analysis by geographic region on the time to first onset of any of the composite events (cardiovascular death, myocardial infarction, and stroke), the primary endpoint, suggested that the treatment effect in North America was different from that in other regions (upper figure in Figure 3-7) and that the results in the US showed high heterogeneity (Section 5.1). However, a subgroup analysis by dose of aspirin, an effect modifier, in the US and non-US showed that the therapeutic effect was reversed in the population receiving a high dose of aspirin, with consistent results both in the US and non-US (lower figure in Figure 3-7).

Figure 3-7 Layer 2: Internal consistency of efficacy case (ticagrelor)

Primary endpoint results by geographic region



Primary endpoint results by aspirin dose in US and non-US



Source: ICH E7 Training Material Module 6

External consistency

Consider whether differences in results between regions or populations, and the impact of potential effect modifiers on treatment response will be observed across multiple data sources. Specifically, if there are multiple studies that have evaluated the same drug in a similar patient population, the effect of candidate effect modifiers evaluated by the internal consistency of Layer 2 on the treatment effect will be examined for consistency across studies. An example of evaluation of external consistency is to confirm the effect of factors affecting the treatment effect in the phase II study, that is, if a candidate effect modifier is identified, by using it as a stratification factor for randomization in the phase III study or by planning a subgroup analysis in advance. Another approach is to evaluate the consistency with the results of studies of drugs of the same class and indication.

Statistical uncertainty

A consideration of uncertainty is an important and indispensable factor in evaluating scientific evidence.

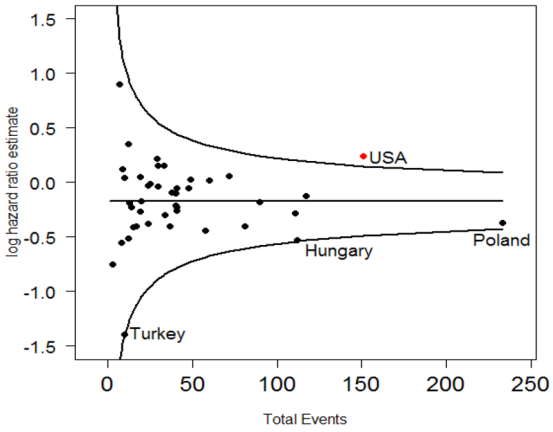
In Layer 2, the accuracy of the estimates will be examined considering how accurate the differences in results observed across regions or populations and the candidate effect modifiers affect the treatment effect.

Case example: Ticagrelor

The figure below shows the PLATO study described in ICH E17 Training Material Module 6 as an example of statistical uncertainty for consistency across regions. In the funnel plot (Figure 3-8), the results of the primary endpoint in the US deviated from the 95% CI, and the p-value of the interaction term between treatment arm and region (US or non-US) was 0.0095, and the probability of ≥ 1.27 in the US when the hazard ratio of the primary endpoint was 0.84 in the overall population was < 0.006, suggesting that there is some systematic factor for the results in the US. However, there is not enough power to test in the US alone, and the possibility that the results in the US are incidental cannot be ruled out.

Figure 3-8 Layer 2: Assessment case of statistically efficacy uncertainty (ticagrelor)

Primary endpoint by participating country - funnel plot



Source: ICH E7 Training Material Module 6

Clinical relevance

In Layer 2, consider whether differences in results observed between regions or populations and the impact of potential effect modifiers on treatment effects are important findings that provide the basis for clinical judgment and treatment policy decisions.

Case example: Ticagrelor

In the PLATO study described in ICH E17 Training Material Module 6, the maintenance dose of aspirin was identified as an effect modifier, showing that the therapeutic effect of ticagrelor is reduced in populations with higher maintenance doses of aspirin. As shown in the internal consistency case, the hazard ratio of ticagrelor relative to the comparator exceeded 1 both in the US and non-US when used in combination with a high dose of aspirin exceeding a maintenance dose of 300 mg, suggesting that ticagrelor does not provide any efficacy benefit relative to the comparator.

3.3.1 Efficacy Assessments

3.3.1.1 Examination of effect modifiers from various perspectives

There are many intrinsic and extrinsic factors that may affect the efficacy and safety of a drug, and there is no uniform method for examination of effect modifiers, so it is necessary to examine the method according to the characteristics of the drug. Examples of approaches for explaining effect modifier based on intrinsic and extrinsic factors are shown in the following sections.

(1) One factor at a time approach

A simple and typical method for identifying the presence of an effect modifier is subgroup analysis, in which a population is divided into subgroups according to some intrinsic or extrinsic ethnic factor and differences in treatment effects are examined. Simple subgroup analyses as well as combined use of the graphical approaches illustrated in the example in Figure 3-9 may be useful. Differences in treatment effects among subgroups suggest that the factor may be an effect modifier. However, this approach can examine the relationship between each factor and the treatment effect but cannot account for the combined effect of multiple factors.

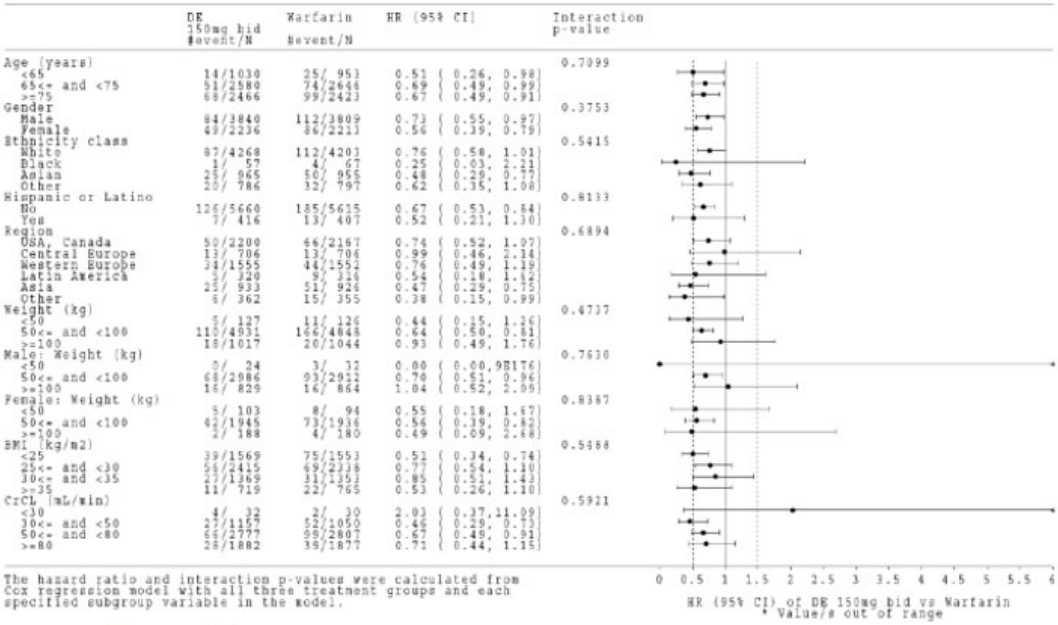
1) Forest plot

One graphical approach to subgroup analysis is a forest plot. Since subgroup analyses are performed to investigate the presence or absence of interactions, the hazard ratio (95% CI) and p-value of the interaction may be provided in addition to summary statistics for each subgroup. As described in 1) Examination of interactions of (3) Approaches using statistical models, the p-value of the interaction should be treated as a reference value only and if the p-value is small, further investigation is needed considering the possibility that the factor is an effect modifier.

Case example: Dabigatran etexilate methanesulfonate (dabigatran)

Figure 3-9 shows a forest plot of the hazard ratios of dabigatran versus warfarin for stroke or systemic embolism from the Phase 3 MRCT (RE-LY). There were no significant interactions between any of the factors and the treatment arm. The incidence of stroke or systemic embolism in the dabigatran group increased with increasing age and creatinine clearance. The trend was similar in the control group.

Figure 3-9 Forest plot: Case example of investigating various effect modifiers for efficacy (dabigatran)



Source: CTD 5.3.5.1-4, Study 1160.26, U[redacted]-3249-01, Table 15.2.3.2: 2

Figure 2.5.4.6.2:1 Hazard ratio and 95% CI for stroke/systemic embolism for BIBR 1048 MS 150-mg BID administration group compared to warfarin administration group by demographic and other baseline characteristics - randomized population

Source: CTD for Dabigatran (2011)

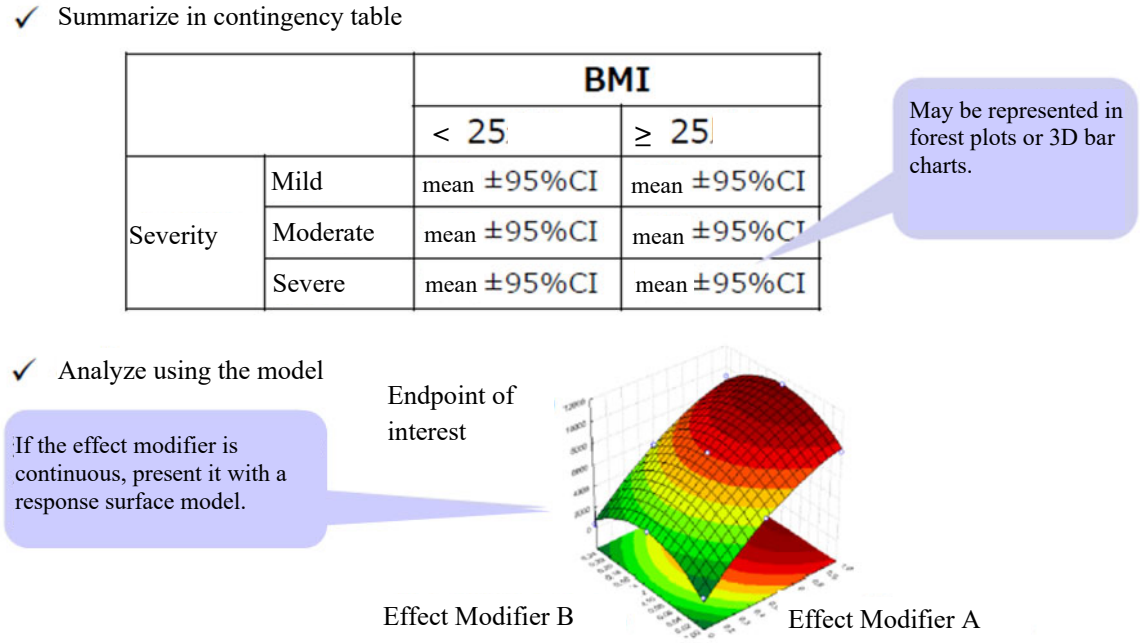
2) Funnel plot

The funnel plot for the evaluation of consistency across regions in Section 3.3.1.2 (Figure 3-15) is a plot of the estimated values on the vertical axis and the number of participants on the horizontal axis for the subgroups to be studied. This plot can also be used to examine intrinsic or extrinsic ethnic factors and is useful for examining treatment effect regularity and outliers between the levels of each factor.

(2) Approach to perform subgroup analysis with a combination of effect modifiers

This approach attempts to account for the combined effects of multiple effect modifiers. Figure 3-10 shows the consideration of the case where severity of disease and BMI are candidate effect modifiers. For example, if the severity of the target disease and BMI are effect modifiers, and if the severity of the disease is divided into three levels (mild, moderate, and severe) and BMI is divided into two levels (< 25, ≥ 25), there will be six subgroups based on the combination of the two effect modifiers. The approach attempts to explain the combined treatment effect of severity and BMI by looking at how the treatment effect differs among these six subgroups. If the endpoint is a continuous variable, a response surface model may also be useful. However, a concern with a larger number of subgroups is that the more poorly informed subgroups with small numbers of participants may be difficult to interpret.

Figure 3-10 Example of subgroup analysis with a combination of effect modifiers (hypothetical case)



Source: 7th ICH E17 Workshop (2022)

(3) Approaches using statistical models

Since statistical models can incorporate multiple effect modifiers simultaneously, it is possible to examine the effect of each effect modifier, and their combined effects, i.e., interactions among effect modifiers can also be investigated. In the subgroup analysis stated above, it is necessary to define a threshold for each effect modifier and divide it into several levels (ordinal categories). In statistical models, effect modifiers can be treated either as ordinal categories or as continuous variables, and interactions can be examined in either case.

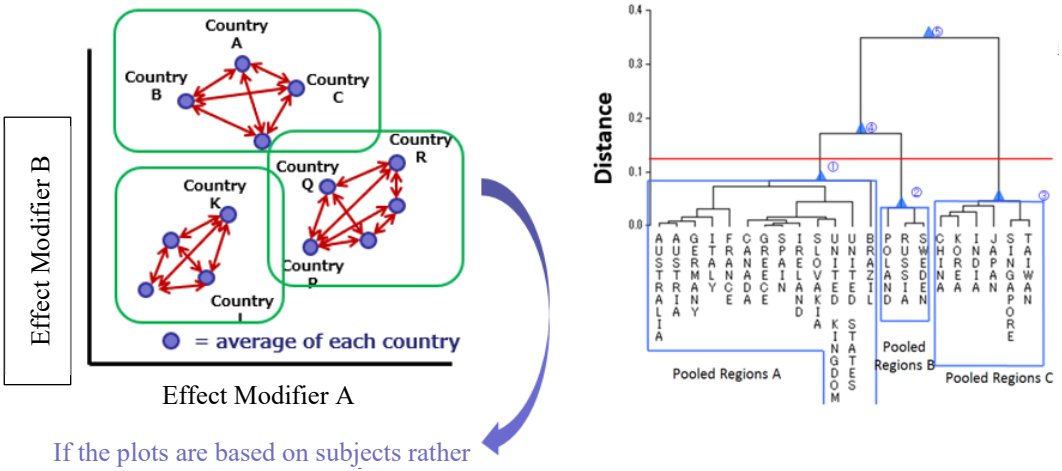
1) Examination of interactions

In this approach, analyses are performed using statistical models including the treatment arm, candidate factors for effect modifier, and interactions between the treatment arm and the candidate factors. Effect modifiers are explored from the test results of the interaction terms. When an interaction test result is significant, it suggests that a factor may be an effect modifier, but the results should be interpreted with caution because the power of the interaction test is generally low and there are no clear criteria for determining the significance level of the test. Clinical validity should also be taken into consideration for evaluation of results instead of judging based on the test results alone.

2) Approaches to pooled analysis with clustering by effect modifiers

This is an approach to explore effect modifiers from the treatment effects and characteristics (distribution of candidate factors) of each group by dividing participants into multiple groups based on the similarities among multiple candidate factors for effect modifier. There are two ways to group: by country or region, or by participant. The grouping by country or region will allow examination of differences in treatment effects among the pooled countries or regions, while the grouping by participant will allow examination of differences in treatment effects among the pooled subpopulations. However, there are no clear criteria for determining the degree of similarity, and if there are many factors to be examined, the characteristics of each group may not be explained.

Figure 3-11 Case example of pooled analysis with clustering by effect modifiers (virtual case)

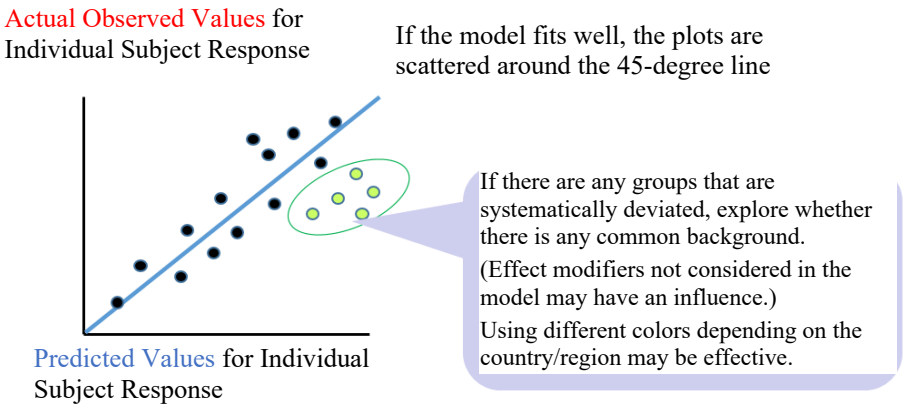


Source: 7th ICH E17 Workshop (2022)

3) Residual plot

A residual plot is a scatter plot used for diagnosis of the constructed model, and can visually capture some regularity, bias of variance, and outlier detection. Figure 3-12is a scatter plot with observed values on the Y-axis and predicted values on the X-axis, which is a Q-Q plot different from the general residual plot (residuals on Y-axis, predicted values on X-axis), but the essence of the plot remains the same although the expression is different. When there is a systematic deviation, effect modifiers may have an influence, but it is not possible to determine from this figure the population to which they belong.

Figure 3-12 Case example of Q-Q plot (virtual example)



Source: 7th ICH E17 Workshop (2022)

4) Galbraith plot

The Galbraith plot is a method of plotting the normalized estimates against the precision (reciprocal of standard error [SE]). It is a scatter plot which is created by first dividing each estimate by its standard error to calculate the standardized estimate or z-statistic and then plotting each z-statistic (vertical axis) versus 1/SE (horizontal axis). This technique is useful for outlier identification; if homogeneous, the data points are distributed within the regression line passing through the origin ±2 standard error (Section 5.1.1 Figure example 1-2).

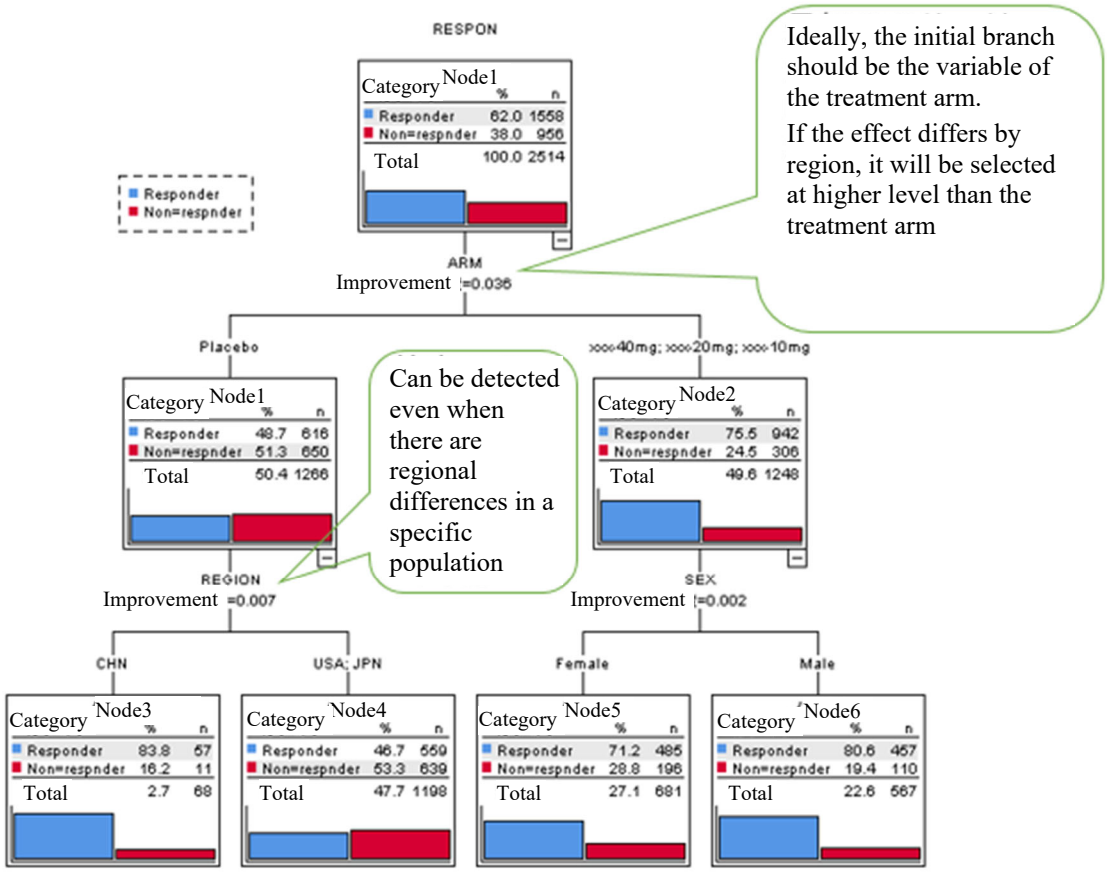
5) CART Analysis

CART is an algorithm of decision tree analysis, which always has two branches. Decision tree analysis is a popular data-mining technique (machine learning) for discrimination and prediction, which has been developed from the AID (Auto Interaction Detection) method and is suitable for the detection of interactions. It is a method to express the factors necessary for decision making and decision criteria in the form of a tree diagram, and the analysis results are automatically selected in the order of the factors that have strong influence on the objective variable and are arranged in a hierarchical order. It helps to easily understand the mutual (hierarchical) relationship between each factor. There are also the following advantages:

- Applicable to non-linear data and data with many explanatory variables (high dimension) without any assumption for data distribution
- Applicable to data with mixed variables (e.g., qualitative vs. quantitative) including missing data
- Visualization of results

However, there are no statistical decision criteria such as the p-value, and attention should be paid since arbitrary judgment is used to determine the level of hierarchy.

Figure 3-13 Case example of CART analysis (hypothetical)



3.3.1.2 Evaluation of consistency across regions

Methods to assess consistency across regions include descriptive summaries, graphical presentations, model-based estimations, and consideration of treatment-by-region interactions as described in ICH E17 Training Material Module 6. The following sections provide examples of approaches to illustrate consistency across regions. As explained in Section 2.1, a consistency evaluation does not mean that the results are consistent among regions. It means that the presence or absence of differences in the results among regions is evaluated multilaterally and systematically, and if there are differences, the factors are identified to examine the impact on benefits and risks. Therefore, if there is a difference among regions, it will be examined whether the difference can be explained by a difference in the effect modifier or its distribution.

(1) One factor at a time approach

Like the analysis by factor, a simple and typical method of analysis by region includes subgroup analysis which is divided by region and examines differences in treatment effect.

1) Forest plot

A forest plot is useful as one of the graphical approaches to evaluate the consistency among regions.

Case example: Finerenone

Figure 3-14 shows a forest plot of the hazard ratio (95% CI) of the incidence of renal failure by country for MRCT with finerenone versus placebo. Finerenone, a mineralocorticoid receptor antagonist, underwent MRCT in patients with chronic kidney disease complicated by type 2 diabetes mellitus in 2 studies. The applicant explained that the occurrence of renal failure varied among countries in both studies and there were no apparent similarities in geographical factors or ethnic factors among countries in which the hazard ratio exceeded 1 within a study or among studies.

Figure 3-14 Forest plot: Case example of consistency evaluation between regions (Finerenone)

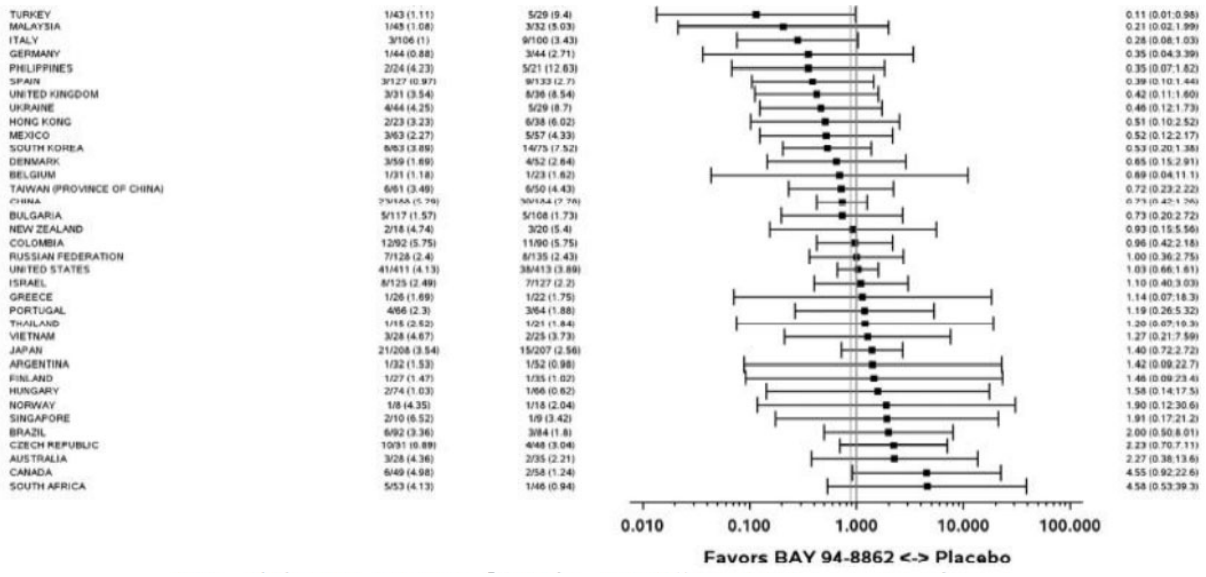


Figure 5. Forest plot of the incidence of renal failure by country in Study 16244 (Cox proportional hazards model with treatment arm as factors in FAS)

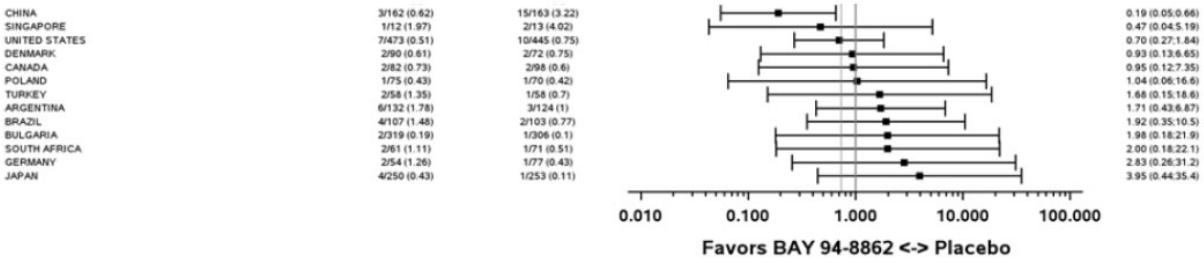


Figure 6. Forest plot of the incidence of renal failure by country in Study 17530 (Cox proportional hazards model with treatment arm as factors in FAS)

Source: Review Report of Finerenone (2022)

2) Funnel plot

The funnel plot plots the estimated values on the vertical axis and the number of participants on the horizontal axis for each subgroup under consideration.

Case example: Dabigatran

Figure 3-15 shows funnel plots of hazard ratios versus warfarin for the occurrence of stroke and systemic embolism by country or region based on dabigatran in a Phase 3 MRCT (RE-LY study) results. With the hazard ratio on the vertical axis and the number of events on the horizontal axis, the 95% CI is shown with a dotted line. The occurrence of composite events in any country was within the range of variability and no geographic factors were found in the countries with hazard ratios greater than 1.

Figure 3-15 Funnel plot: Case example of consistency evaluation of efficacy across regions (dabigatran)

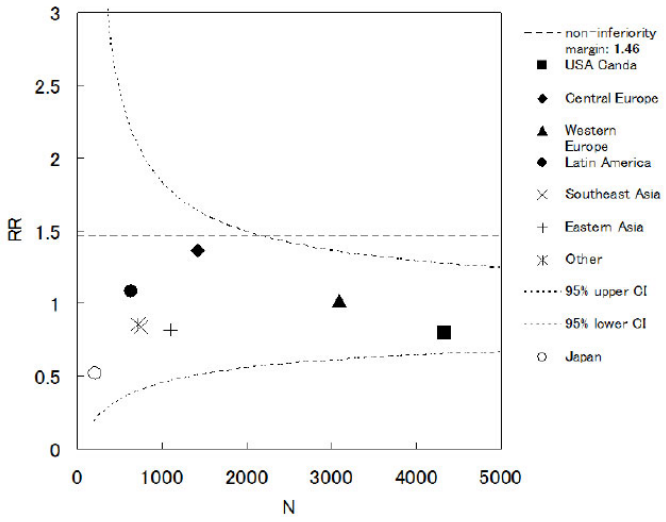


Figure 3-3 Hazard ratio of stroke and systemic obstruction by region (110 mg group vs. warfarin group) 10

Source: The practice of 3-layer approach that is the methods for examining ethnic factors in multi-regional clinical trials (Japan Pharmaceutical Manufacturers Association, 2018)

(2) Model-based estimation

As for methods for estimating the treatment effect by country/region based on models, the following 3 methods have already been introduced in "The practice of 3-layer approach that is the methods for examining ethnic factors in multi-regional clinical trials" (Japan Pharmaceutical Manufacturers Association, 2018).

1) Use of shrinkage estimator

Two methods used to estimate the treatment effect in each region are the empirical shrinkage estimator proposed by Quan et al. to evaluate the consistency among regions and the James-Stein type shrinkage estimator.

2) Tree approach considering regional similarities

To correct the fact that the shrinkage estimator proposed by Quan et al. does not consider the similarities between regions, Guo et al. proposed an analysis method in a Bayesian framework considering the similarities by the human development index (HDI) between countries, based on the HDI recommended by the multi-regional clinical trial group PhRMA.

3) Estimation of treatment effects by region by applying standardized methods

The standardized method using the theory of stratified analysis can be applied to estimate the treatment effect by region if ethnic factors affecting the treatment effect have been identified. Using the standardized method, it is possible to obtain estimates for each region using all the data in a relatively simple manner.

Each of these 3 methods has advantages and disadvantages as shown in Table 3-1 and it is necessary to select the appropriate method. For details of these methods, refer to "The practice of 3-layer approach that is the methods for examining ethnic factors in multi-regional clinical trials" (Japan Pharmaceutical Manufacturers Association, 2018).

Table 3-1 Advantages and disadvantages of each approach for treatment effect in each country/region based on modeling

	Advantage	Disadvantage
Shrinkage estimator	It is easy to show consistency because it is reduced toward the overall estimate	Does not reflect similarities or differences in ethnic factors (depends only on sample size and variability)
Tree approach	Reflected overall similarity across regions	Calculation is complicated The similarity reflected is general and does not directly reflect the impact of factors unique to the drug.
Standardized method	The calculation method is easy to understand and the results are easy to interpret	Influencing factors should be identified and collected as data Many influencing factors cannot be considered.

(3) Examination of interactions between treatment and region

As in the case example of the forest plot to examine effect modifiers from various perspectives in Section 3.3.1.1 (Figure 3-9), subgroup analyses by country or region are conducted for the purpose of examining the presence or absence of interactions, and p-values for interactions may

be presented. However, as described above, p-values for interactions should be handled as reference values only. Some of the possible causes of the difference in the treatment effect between regions include problems in the conduct of the study such as deviations from GCP and the protocol. After those possibilities are ruled out, it is necessary to investigate whether other effect modifiers can explain the difference in the treatment effect between regions when the p-value of the interaction is small.

3.3.2 Safety Assessments

As described in ICH M4 E(r2), in order to consider treatment and management based on intrinsic or extrinsic ethnic factors for each patient, it is also important to evaluate safety in Layer 2 focusing on the investigation of effect modifiers. Safety evaluations includes adverse events (AEs), laboratory data, vital signs, electrocardiograms (ECGs), and other findings, and it is important to assess the pattern and relevance of a wide variety of related items. For adverse events that occur relatively frequently, the following approaches may also be useful. However, it is difficult to evaluate low-incidence adverse events including serious adverse events as a group, and it is necessary to identify risk factors after evaluating the presence or absence of common factors among the participants based on the incidence patterns and background factors of individual participants. In addition, the safety evaluation should be presented comprehensively and logically, including not only results obtained from clinical studies but also non-clinical studies, drug interaction studies, population pharmacokinetic data and, if available, post-marketing information and information on similar drugs.

ICH M4E(r2)	2.7.4.5 Safety in Special Groups and Situations 2.7.4.5. 1 Intrinsic Factors This section should summarize safety data pertinent to individualized therapy or patient management based on demographic and other factors defined as “intrinsic ethnic factors” in ICH E5. These factors include age, sex, height, weight, lean body mass, genetic polymorphism, body composition, other illness, and organ dysfunction. Safety in the pediatric population should be routinely analyzed in applications for a proposed indication for children. Analysis of the impact of such factors on safety outcomes should have been presented in other sections but should be summarized here, together with pertinent PK or other information, e.g., in patients with renal or hepatic disease. If a sufficiently large number of subjects with a given co-morbid condition such as hypertension, heart disease, or diabetes, was enrolled, analyzes should be carried out to assess whether the co-morbid condition affected the safety of the drug under study. Cross reference should be made to the tables or description of adverse events when analyses of such sub-groups has been carried out. 2.7.4.5. 2 Extrinsic Factors This section should summarize safety data pertinent to individualized therapy or patient management based on factors defined as “extrinsic ethnic factors” in ICH E5. These are factors associated with the patient environment. Examples are the medical environment, use of other drugs (see 2.7.4.5.3, Drug Interactions), use of tobacco, use of alcohol, and food habits. For example, if a potential interaction with alcohol is suggested by the metabolic profile, by the results of studies, by post-marketing experience, or by information on similar drugs, information should be provided here.

3.3.2.1 Examination of effect modifiers from various perspectives

(1) One factor at a time approach

1) Frequency distribution table

In safety evaluation, as with the efficacy assessments, a simple and typical approach involves subgroup analyses that examine differences in the frequency of adverse events and other safety findings (laboratory test values, vital signs, ECG) among subgroups divided by some intrinsic or extrinsic ethnic factor.

Case example: Finerenone

In the review report of finerenone, PMDA asked the applicant to explain the necessity of raising awareness because the risk of decreased blood pressure was assumed from the mechanism of action of this drug. The applicant explained it including the occurrence of decreased blood pressure-related events by baseline systolic blood pressure as follows.

Table 3-2 Frequency distribution table: Case example of investigating effect modifiers from various perspectives of safety (finerenone)

Table 56. Incidence of blood pressure-related events by baseline systolic blood pressure in the overall population (Study 16244: Safety Analysis Set)

	Systolic blood pressure < 100 mg		Systolic blood pressure ≥ 100 to < 130 mmHg		Systolic blood pressure ≥ 130 to < 160 mmHg		Systolic blood pressure ≥ 160 mmHg	
	Placebo (N=16)	Drug (N=9)	Placebo (N=761)	Drug (N=779)	Placebo (N=1916)	Drug (N=1897)	Placebo (N=138)	Drug (N=141)
Any hypotension-related event[a]	12.5 (2)	11.1 (1)	6.0 (46)	8.9 (69)	3.0 (57)	3.9 (74)	2.2 (3)	3.5 (5)
Dose discontinuation	0 (0)	0 (0)	0 (0)	0.1 (1)	0 (0)	0 (0)	0 (0)	0 (0)
Serious	0 (0)	0 (0)	0.3 (2)	0.5 (4)	0.2 (4)	0.4 (7)	0 (0)	0 (0)
Hospitalization	0 (0)	0 (0)	0.3 (2)	0.4 (3)	0.2 (4)	0.4 (7)	0 (0)	0 (0)
Death	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

Incidence% (N)

a : MedDRA PT Blood pressure decreased, Hypotension, Orthostatic hypotension

Table 57. Incidence of blood pressure-related events by baseline systolic blood pressure in the overall population (Study 17530: Safety Analysis Set)

	Systolic blood pressure < 100 mg		Systolic blood pressure ≥ 100 to < 130 mmHg		Systolic blood pressure ≥ 130 to < 160 mmHg		Systolic blood pressure ≥ 160 mmHg	
	Placebo (N=26)	Drug (N=14)	Placebo (N=1168)	Drug (N=1173)	Placebo (N=2351)	Drug (N=2389)	Placebo (N=113)	Drug (N=107)
Any hypotension-related event[a]	34.6 (9)	28.6 (4)	4.5 (53)	7.5 (88)	2.0 (48)	3.8 (91)	4.4 (5)	5.6 (6)
Dose discontinuation	0 (0)	7.1 (1)	0 (0)	0 (0)	< 0.1 (1)	< 0.1 (1)	0 (0)	0 (0)
Serious	0 (0)	0 (0)	0.2 (2)	0 (0)	0 (0)	0.1 (3)	1.8 (2)	0 (0)
Hospitalization	0 (0)	0 (0)	0.2 (2)	0 (0)	0 (0)	0.1 (3)	1.8 (2)	0 (0)
Death	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

Incidence% (N)

a : MedDRA PT Blood pressure decreased, Hypotension, Orthostatic hypotension

Source: Review Report of Finerenone (2022)

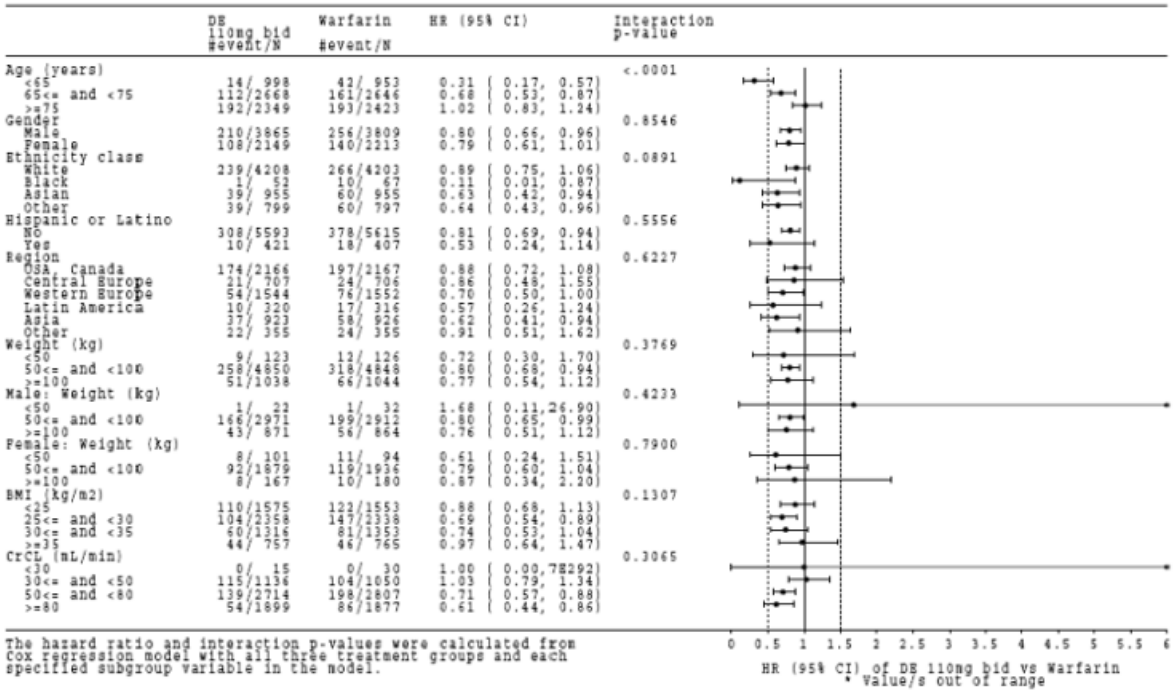
2) Forest plot

Forest plots may be useful as a graphical approach for safety evaluation as well as for efficacy evaluation.

Case example: Dabigatran

Figure 3-16 shows a forest plot of the hazard ratios of dabigatran versus warfarin for the occurrence of major bleeding in a Phase 3 MRCT (RE-LY). A forest plot is useful tool to evaluate the association of intrinsic or extrinsic ethnic factors with the occurrence of specific adverse events. However, as described above, the p-value of the interaction should be handled as a reference value only, and if the p-value is small, further investigation is needed considering the possibility that the factor is an effect modifier.

Figure 3-16 Forest plot: Case example of evaluating effect modifiers from various safety perspectives (dabigatran)



Source: CTD2.7.4, Appendix 1, Figure 2.1.1.1.26

Figure 2.5.5.1.7:1 Comparison of hazard ratio and 95% CI for major bleeding by demographic characteristic in Study 1160.26 (BIBR 1048 MS 110-mg BID group vs warfarin group), randomized set

Source: CTD for Dabigatran (2011)

3) Attributable risk ratio (ARR)

One way to identify whether one subgroup differs from another is to calculate the ARR for differences between subgroups (The Science of Drug Safety, 2012). Although ARR is useful

for high-incidence events and other findings that are high-incidence and potentially drug related (e.g., laboratory values, vital signs), the analysis is not possible for low-incidence events.

ARR = (incidence of adverse events with investigational drug X – incidence of adverse events with comparator X)／(incidence of adverse events with investigational drug Y – incidence of adverse events with comparator Y)

X and Y represent subgroups (e.g., X = ≥ 65 years; Y = < 65 years)

(2) Approaches using statistical models

To explore factors influencing the occurrence of specific adverse events for a safety evaluation, it may be useful to explore the factors with strong influence by introducing a statistical model and using a variable selection method.

Case example: Enforzumab vedotin (genetical recombination)

In the review report of enforzumab vedotin, which was approved for the treatment of unresectable urothelial cancer that has progressed after cancer chemotherapy, PMDA asked for the applicant's view on the mechanism of development of skin disorder and risk factors. The applicant explained that, in addition to general subgroup analyses, the relationship between the skin disorder and 22 potential risk factors was investigated based on variable selection by a multivariate logistic regression model and that the following risk factors were identified: non-white race, high hemoglobin level, and small tumor diameter (odds ratio [95% CI]: (1) non-white race 1.947 [1.182 3.210], (2) hemoglobin level increased by 1 g/dL 1.024 [1.009, 1.039], (3) tumor diameter increased by 1 mm 0.991 [0.984, 0.997]).
Source: Review Report of Enforzumab Vedotin (Genetical Recombination) (2021)

As multiple effect modifiers can be incorporated into the statistical model simultaneously, the influence of each effect modifier and their composite influences, i.e., interaction among effect modifiers, can also be investigated, and thus may also be used for the safety evaluation.

Case example: Dabigatran

In the review report of dabigatran, PMDA asked the applicant to justify age ≥ 75 years, concomitant use of P-glycoprotein inhibitors, and history of gastrointestinal hemorrhage as high-risk patients for hemorrhage. As 1 case, the applicant explained the following 2 points as the rationale for the age of ≥ 75 years.

- In Study 1160.26, a stratified analysis of the incidence of bleeding between patients aged ≥ 75 years and those aged <75 years showed that the annual event rates of major bleeding and all bleeding were higher in patients aged ≥ 75 years in both this drug and warfarin groups.
- Analysis using a Cox regression model (model with age, baseline creatinine clearance, sex, and concomitant use of aspirin as covariates and interaction of treatment arm by each covariate) showed that the hazard ratio of this drug to warfarin tended to increase with

increasing age, with the hazard ratio exceeding 1 at the age of >85 years in the 110-mg group for this drug and at the age of >75 years in the 150-mg group for this drug.

Figure 3-17 Approaches using statistical models: Case example of considering effect modifiers from various perspectives of safety (dabigatran)

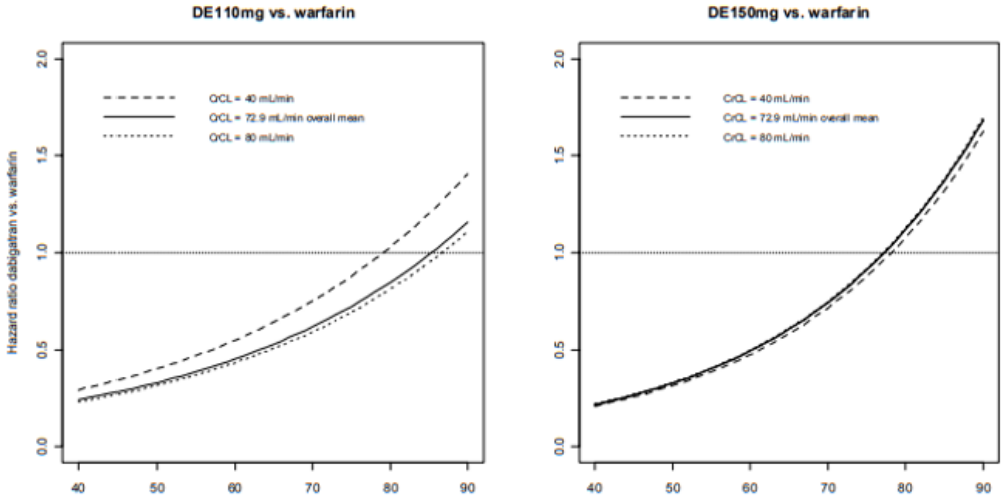


Figure 2.7.4.2.2.1.1.2: 8 Hazard ratio of major bleeding for age at a specified creatinine clearance value in Study 1160.26 (randomized population)

Source: CTD2.7.4, Appendix 1, Table 2.1.1.1.7.3

Source: CTD for Dabigatran (2011)

3.3.2.2 Evaluation of consistency across regions

(1) One factor at a time approach

Like the analysis by factor, a simple and typical method of analysis by region includes subgroup analysis by region to examine differences in treatment effect. Graphical approaches may also be useful for investigating the occurrence of specific adverse events.

Case example: Dabigatran

In the dabigatran application dossier (CTD 2.7.4), the hazard ratio for major bleeding by region, the number of participants, and the corresponding 95% CI are presented in the following frequency distribution tables (Table 3-3) and funnel plot (Figure 3-18).

Table 3-3 Frequency distribution table: Evaluation of consistency of safety between regions (dabigatran)

Table 2.7.4.2.3.2.1: 3 Annual major bleeding event rate and hazard ratio, and CI in each study region

Region/Country	Subjects with event / Number of randomized subjects (Yearly event rate %)			Hazard ratio (95% CI)	
	DE 110 bid	DE 150 bid	Warfarin	DE 110 bid vs Warfarin	DE 150 bid vs Warfarin
Total	318 / 6015 (2.67)	375 / 6076 (3.11)	396 / 6022 (3.36)	0.79 (0.68, 0.92)	0.93 (0.81, 1.07)
North America	174 / 2166 (3.92)	208 / 2200 (4.61)	197 / 2167 (4.45)	0.88 (0.72, 1.08)	1.04 (0.86, 1.27)
Latin America	10 / 320 (1.82)	15 / 320 (2.73)	17 / 316 (3.18)	0.57 (0.26, 1.24)	0.86 (0.43, 1.72)
Central Europe	21 / 707 (1.53)	23 / 706 (1.69)	24 / 706 (1.77)	0.86 (0.48, 1.55)	0.96 (0.54, 1.69)
Western Europe	54 / 1544 (1.74)	65 / 1555 (2.07)	76 / 1552 (2.46)	0.70 (0.50, 1.00)	0.84 (0.61, 1.17)
Southeast Asia	11 / 378 (1.60)	15 / 381 (2.09)	24 / 375 (3.56)	0.44 (0.22, 0.91)	0.59 (0.31, 1.13)
Eastern Asia	26 / 545 (2.43)	21 / 552 (1.95)	34 / 551 (3.22)	0.74 (0.45, 1.24)	0.60 (0.35, 1.03)
Other	22 / 355 (3.30)	28 / 362 (4.13)	24 / 355 (3.64)	0.91 (0.51, 1.62)	1.15 (0.67, 1.98)
Japan	8 / 107 (5.53)	5 / 111 (3.33)	5 / 108 (3.31)	1.68 (0.55, 5.15)	1.02 (0.29, 3.51)

Source : CTD 2.7.4, Appendix 1, Table 2.1.1.1.1.3, Table 2.1.1.1.1.14, Table 2.1.1.1.1.24, Figure 2.1.1.1.1.26, Figure 2.1.1.1.1.27; Appendix 3, Table 15.3.5.3:3, Table 15.3.2.1:9; Appendix 4, Table 15.3.2.1:2, Table 15.3.2.1:9
CTD 5.3.5.1-4, Study 160.26, U[redacted]-3249-01, Table 15.3.5.3: 3, Table 15.3.5.3: 6

Source: CTD for Dabigatran (2011)

Figure 3-18 Funnel plot: Case example of consistency evaluation of safety between regions (dabigatran)

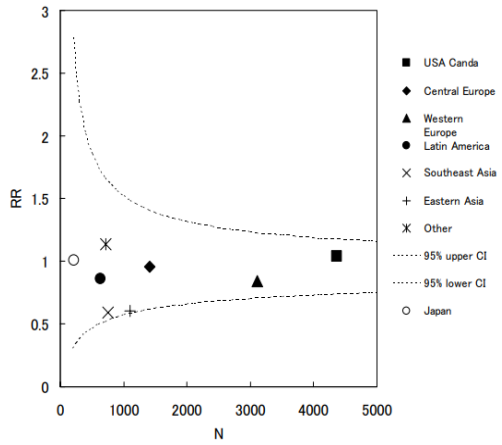


Figure 2.7.4.2.3.2.1: 3
Hazard ratio and number of subjects with major bleeding in each study region (DE 150 bid vs Warfarin)

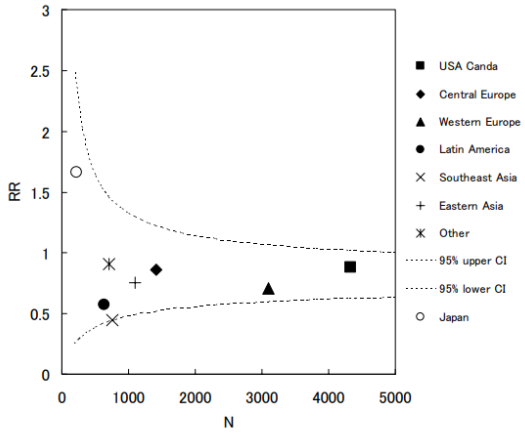


Figure 2.7.4.2.3.2.1: 2
Hazard ratio and number of subjects with major bleeding in each study region (DE 110 bid vs Warfarin)

RR Relative Risk N : Number of subjects
Note: The 95% CI was calculated by adjusting the 95% CI (logarithmic value) calculated from all patient data with the square root of the number of subjects in each participating region.
Source : CTD 2.7.4, Appendix 1. Table 2.1.1.1.3, Table 2.1.1.1.1.24; Appendix3, Table 15.3.5.3:3; Appendix 4, Table 1.53.2.1: 2 : CTD 5.3.5.1-4, Study 1160.26, U[redacted]-3249-01, Table 15.3.5.3: 3

Source: CTD for Dabigatran (2011)

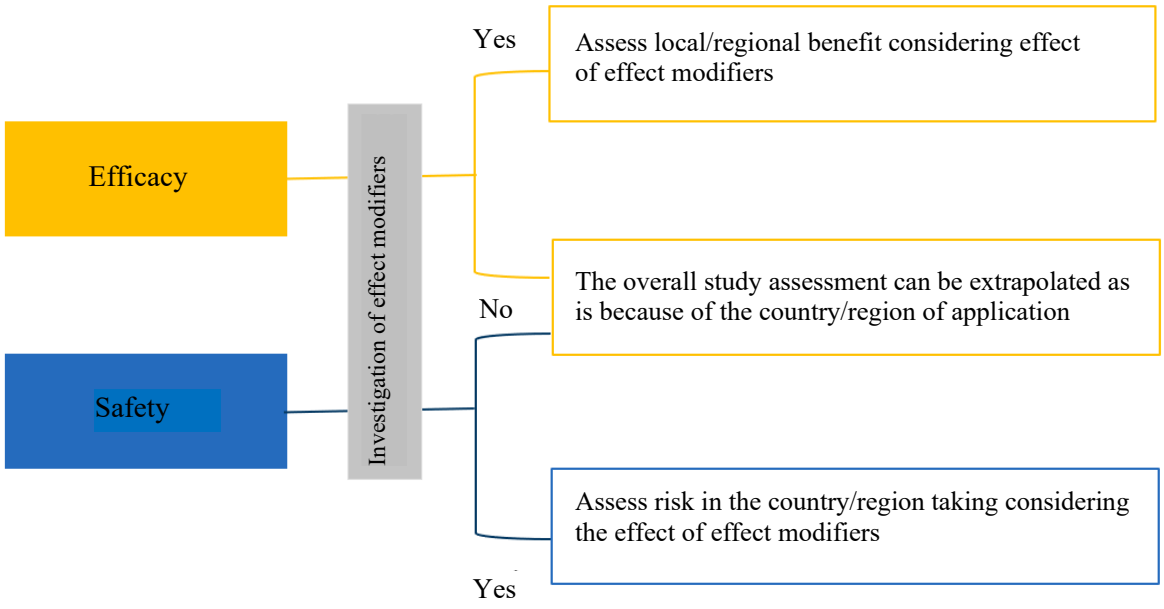
(2) Examination of interactions between treatment and region

As in the case example of the forest plot to examine effect modifiers from various perspectives in Section 3.3.1.2 (Figure 3-14), subgroup analyses by country or region are conducted for the purpose of examining the presence or absence of interactions, and p-values for interactions may be presented. However, as described above, the p-value of the interaction should be handled as

a reference value only, and if the p-value is small, it is necessary to examine whether other effect modifiers can explain the difference in treatment effect between regions.

3.4 Layer 3: Benefit-Risk Assessment in the Country/Region of Application

Layer 3 is a hierarchy for evaluating the usefulness of the investigational drug from both the efficacy and safety aspects based on the results of Layer 1 and Layer 2, and finally evaluating the benefit-risk balance in the country/region of application. The evaluation up to Layer 2 is a comprehensive evaluation of what findings have been obtained in the MRCT as a whole and whether the findings are consistent with the external information of the MRCT (findings from previous studies, other MRCT findings obtained at the same time as the MRCT), and is an evaluation that can be shared widely, not only in the countries that actually participate in MRCT but also worldwide. Based on these assessments, Layer 3 explains what benefits and risks can be expected in each country or region. If the Layer 2 evaluation does not identify any clinically relevant effect modifiers or candidates for both efficacy and safety, the evaluation of the entire study will be considered extrapolable as is, because of the country/region of application. Once an effect modifier or candidate is identified for efficacy or safety, the benefit-risk assessment in the country/region of application should consider the influence of the effect modifier.



The following sections describe considerations that should be kept in mind when assessing the benefit-risk for the country/region of application, where an effect modifier has been identified.

3.4.1 Benefit-Risk Assessment in the Country/Region of Application When Effect Modifiers Have Been Identified

Once an effect modifier is identified, the distribution of effect modifiers in the country/region of application will be used to estimate the efficacy and safety of investigational drug. The source

of the distribution of effect modifiers refers to clinical trial data and external information such as the literature and local statistical data to characterize the patient population in the country/region by effect modifiers. An example where effect modifiers have been identified is shown below, where the patient population in the country/region of application is limited by effect modifiers based on the PLATO study of ticagrelor. The details are shown in Section 5.1.

In addition, when feasible, it may also be useful to estimate the treatment effect in the patient population of the country/region and to assess its benefit-risk based on predictive models using effect modifiers.

Case example: Ticagrelor

The PLATO study, provided in ICH E17 Training Material Module 6, found that the therapeutic effect of ticagrelor was lower in the population with higher maintenance doses of aspirin. This has led to including precautions regarding maintenance doses of aspirin in local package inserts.

- FDA: Maintenance doses of aspirin above 100 mg reduce the effectiveness of ticagrelor and should be avoided.
- EMA: Special Warning "Concomitant administration of ticagrelor and aspirin at a maintenance dose of 300 mg or higher is not recommended."
- NMPA: Precautions stating that the maintenance dose of aspirin is 75-100 mg
- PMDA: Precautions for Dosage and Administration "This drug should be administered in combination with aspirin (81-100 mg/day as the maintenance dose)."

4 CTD

4.1 Contents to be Considered in the CTD

This section proposes a basic principle for preparing the CTD based on the principles of ICH E17. In order to prepare the CTD based on the principles of ICH E17, it was considered necessary to add information different from the conventional CTD and change the basic concept.

First, it is necessary to enhance the examination of important effect modifiers affecting the treatment effect. Specifically, as described in Section 3, the results of exploring potential effect modifiers not only in the country/region of application but also in the overall study should be presented to help predict and understand the treatment effect of the drug.

Second, since the information is required to determine whether the product can be approved in the country/region of application, it is necessary to present the benefit-risk assessment in the country/region of application based on the evaluation of the overall study. However, this does not mean that it is necessary to evaluate consistency by comparing the MRCT results between the overall population and the population in the country/region of application, but it would be sufficient to demonstrate the generalizability to the country/region of application, that the population in the country/region of application does not respond differently from other countries/regions, and that even when effect modifiers are found, the benefit-risk is favorable if it is appropriately managed.

As stated in Section 2.1, a conclusion cannot be drawn based on a single evaluation (whether the results in the population of the country/region of the applicant showed the same tendency as the results in the overall population for the primary/secondary endpoints of a MRCT as described above), but it is important to conduct a multifaceted and structured evaluation of consistency in the MRCT results. Therefore, it is considered necessary to discuss the overall MRCT results based on the five perspectives in accordance with the principles of ICH E17.

To prepare the CTD with the above elements, it is recommended to use a 3-layer approach, as described in Section 3. The 3-layer approach is useful for incorporating these elements and preparing systematic and easy-to-understand CTDs.

Based on such basic principles, the contents to be described and the sections of CTD are proposed below.

4.2 Contents and Sections of the CTD

As described in Section 4.1, it is recommended to show the results and discussion based on the 3-layer approach to prepare the CTD based on the principles of ICH E17. Refer to Section 3 for the outline of the 3-layer approach.

When considering the overall MRCT results as well as the evaluations of effect modifiers and consistency across regions, it is recommended to examine them based on the five perspectives shown in ICH E17. Refer to Section 2 for an overview of Section ICH E17.

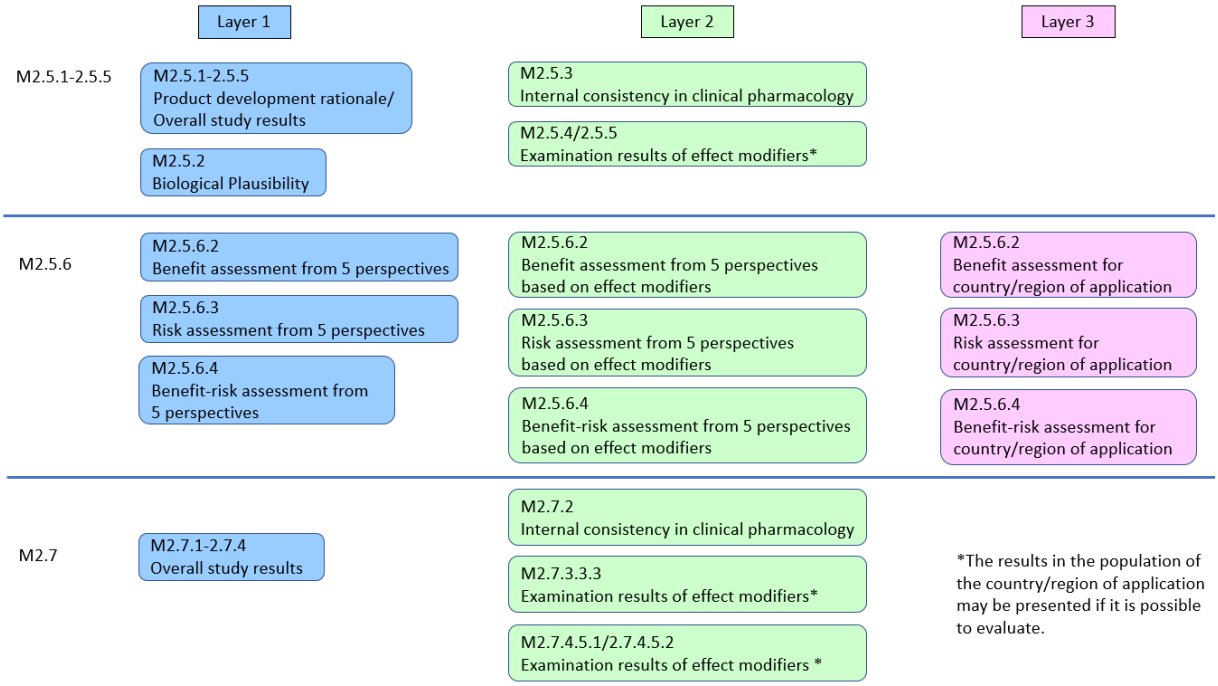
Basic elements for preparing the CTD based on the principles in ICH E17 and the recommended sections in the CTD are shown by layer in Figure 4-1.

For ease of understanding the overview of CTD based on the principles in ICH E17, this section and Figure 4-1 focus on the elements that particularly need to be addressed. The overall study results that have been generally described and components that are not particularly changed are simplified.

First, in Layer 1, the overall MRCT results, discussion from the five perspectives, and the benefit-risk assessment based on the overall study results are described. In Layer 2, the results of examination on effect modifiers, the discussion from the five perspectives based on them, and the benefit-risk assessment are described as the evaluation of factors affecting the overall MRCT results and the consistency evaluation across regions. Finally, in Layer 3, the benefit-risk assessment for the country/region of application based on the benefit-risk assessment of the overall MRCT results through Layer 2 is described.

An outline of the contents to be described in each layer is described in Section 4.2.1 to Section 4.2.3.

Figure 4-1 Basic components to prepare the CTD based on the principles of ICH E17



Detailed explanation for M2.5.6 is shown in Section 4.3, as it particularly needs to be addressed to prepare the CTD based on the basic principles proposed in this document. In addition, examples of CTDs prepared according to the basic concept in this section are shown in Section 5.

4.2.1 Evaluation of Overall Study Results (Layer 1)

This section describes the evaluation of overall MRCT results which should be presented as Layer 1.

In Layer 1, the product development rationale and overall MRCT results are presented. It has been described in the CTD in accordance with ICH M4 E(r2), and no additional actions are recommended in this document. Overall study results are described in Module 2.5.1 through Module 2.5.6 and Module 2.7.1 through Module 2.7.4 as Layer 1 components as described in ICH M4 E(r2).

However, in order to prepare the CTD based on the principles of ICH E17, it is recommended to perform a benefit-risk assessment based on the multifaceted and structured evaluation of overall MRCT results using the five perspectives described in ICH E17. Details of the five perspectives in ICH E17 that should be discussed for the overall study results are shown in Section 3.2, Section 4.3.2.1, and Section 4.3.3.1.

It is recommended to present a discussion of the overall study results based on the five perspectives in ICH E17 in M2.5.6.2, M2.5.6.3, and M2.5.6.4. The source data/information used for the discussion should be presented in the relevant sections from M2.5.1 to M2.5.6 and from M2.7.1 to M2.7.4.

4.2.2 Examination of Effect Modifiers and Consistency Evaluation Across Regions (Layer 2)

This section describes the examination of effect modifiers for the overall study results and the consistency evaluation across regions, which should be presented as Layer 2.

Layer 2 is the most important element in the CTD based on the principles of ICH E17. As described in Section 3.3, the objectives of Layer 2 are to explore effect modifiers or their candidates, and if any inconsistencies among countries/regions or populations are observed, they are explained based on effect modifiers or differences in distribution. Therefore, the main contents to be presented in Layer 2 are subgroup analyses and other results which were conducted to explore effect modifiers or to evaluate the consistency across regions, discussion of effect modifiers from the five perspectives in ICH E17, and the benefit-risk assessment based on the results.

Because the information up to Layer 2 can be shared globally, it is important to discuss the causes even if any inconsistencies are observed in populations outside the country/region of application or in populations considered not directly related to the country/region of application. In addition, it should be considered to what extent effect modifiers should be examined for each drug/application in Layer 2. It is not always good to do more exploration, and it is recommended to identify the items needed to consider Layer 3. This document describes the case of applications in the US, EU, Japan, and China.

Examples of analysis methods for exploring effect modifiers and evaluating the consistency across regions are shown in Section 3.3. It is recommended to present the subgroup analyses and other results which were conducted to explore effect modifiers or to evaluate the consistency across regions in M2.5.4 and M2.5.5 and M2.7.3.3.3, M2.7.4.5.1, and M2.7.4.5.2.

It is recommended to conduct multifaceted and structured evaluation of the results obtained from exploration of effect modifiers or the consistency evaluation across regions based on the five perspectives described in ICH E17. The five perspectives in ICH E17 from which the examination results of effect modifiers should be discussed are shown in Section 3.3, Section 4.3.2.2, and Section 4.3.3.2.

It is recommended to refer to Table 4-1 for the direction of discussion on intrinsic and extrinsic factors (effect modifiers) that influence treatment effects and inconsistencies among countries/regions.

If no effect modifiers that may significantly affect the study results are identified, and no inconsistencies among countries/regions are observed from the examinations results of effect modifiers and discussion from the five perspectives, it can be concluded that there is no evidence to deny the applicability of the overall study results to the country/region of application based on the examination results to date. Therefore, the overall MRCT results can be applied to the country/region of application.

If an effect modifier that may significantly affect the study results is identified, the efficacy or safety of the study product in each region/country should be evaluated in consideration of the influence, irrespective of regional differences in the distribution of factors. Points to be considered for effect modifiers are detailed in Section 3.3, Section 4.3.2.2, and Section 4.3.3.2.

Table 4-1 Direction of discussion on intrinsic and extrinsic factors (effect modifiers) and inconsistencies among countries/regions that influence treatment effects

		Intrinsic and extrinsic factors (effect modifiers)	
		No	Yes
Inconsistency among countries/ regions	No	[Case A] The overall results are robust and the overall conclusions can be applied to all countries/regions.	[Case B] The resulting influence is considered to be globally common and a consistent approach for effect modifiers can be applied to all countries/regions.
	Yes	[Case C] Factors that have not yet been examined should be further examined (other than countries/regions) to identify any confounding factors between countries/regions and intrinsic/extrinsic factors. 1. If new intrinsic or extrinsic factors (effect modifiers) are identified in further examinations, take the same steps as in Case B. 2. If no effect modifiers that affect the overall results are identified in further examinations, other factors (e.g., statistical	[Case D] Since confounding between countries/regions and intrinsic/extrinsic factors may be present, it should be examined whether any influence identified among countries/regions can be explained through other intrinsic/extrinsic factors. 1. If the influence identified among countries/regions can be explained through other intrinsic and extrinsic factors, take the same step as in Case B. 2. If the influence identified among countries/regions cannot be

		Intrinsic and extrinsic factors (effect modifiers)	
		No	Yes
		uncertainty or study conduct conditions) should be considered. This should also be considered in the benefit-risk assessment in the country/region where the influence was observed under Layer 3.	explained through other intrinsic/extrinsic factors, other factors (e.g., statistical uncertainty or study conduct conditions) should be considered as in Case C. This should also be considered in the benefit-risk assessment in the country/region where the influence was observed under Layer 3.

Source: Yoshida et al. 2015 Modified

It is recommended to present a discussion of effect modifiers from the five perspectives based on the subgroup analysis results and the benefit-risk assessment based on it in M2.5.6.2, M2.5.6.3, and M2.5.6.4.

In addition, if a sufficient number of participants evaluable only in a country/region of application can be included in a MRCT, it may be useful to provide the results in the population of the country/region of application as part of Layer 2. However, it is also important to evaluate consistency of the overall study (i.e., evaluation of results based on the five perspectives) in accordance with the basic principles of ICH E17, rather than to evaluate consistency of the results between the overall population and the population of the country/region of application. Therefore, it is recommended to present the results in the population of the country/region of application as a part of the subgroup analysis of Layer 2.

If the subgroup analysis results in the population of the country/region of application are included, it is recommended to present it in M2.5.4 and M2.5.5, M2.7.3.3.3, and M2.7.4.5.

4.2.3 Benefit-Risk Assessment by Region (Layer 3)

This section describes the benefit-risk assessment in the country/region of application to be presented as Layer 3. This is proposed for the preparation of the CTD for an approval application in the US, EU, Japan, and China.

The purpose of Layer 3 is to discuss or estimate the benefit-risk in a country/region based on the results/considerations in Layer 1 and Layer 2, and to express the characteristics of a country/region or population for which treatment effects are to be estimated using effect modifiers. Therefore, the major content to be presented for Layer 3 is the benefit-risk assessment in the country/region of application, which is based on the benefit-risk assessment (including strength of evidence, limitations, and uncertainties) in the overall study in Layer 1 and Layer 2. Specifically, inconsistencies or imbalances in intrinsic or extrinsic patient background factors between the country/region of application and other countries/regions will be discussed in consideration of the evaluation of effect modifiers in Layer 2. If influences of differences in intrinsic or extrinsic patient background factors on the efficacy and safety have been observed in the results of Layer 1 and Layer 2, the influences of differences or imbalances among regions in these factors on efficacy and safety will be evaluated. Points to be considered

for the benefit-risk assessment in the country/region of application are detailed in Section 3.4 and Section 4.3.2.3.

Sources of information may be derived not only from the data collected during the development program, but also from published literature, local or global statistical information, and real-world data.

The discussion based on the information obtained by the time of application is mainly about patient background factors, and the discussion of environmental factors such as the medical environment in the country/region of application is likely to be limited in many cases. Therefore, a full-scale discussion on environmental factors is expected to be conducted after marketing.

For Layer 3, it is recommended to conduct the benefit-risk assessment in the country/region of application based on the results of Layer 1 and Layer 2. For Layer 3, the benefit-risk assessment for the country/region of application is basically presented, but the assessment in other countries/regions may be helpful for the evaluation in the country/region of application. In this document, even if subgroup analyses in the country/region of application are performed in a MRCT, the results will be treated as Layer 2 (part of subgroup analyses), not Layer 3, as described above.

It is recommended to present the benefit-risk assessment for the country/region of application in M2.5.6.2, M2.5.6.3, and M2.5.6.4.

4.3 Items to be Considered and Described in Each Section of M2.5.6 (Conclusion of Benefits and Risks) and Its Source Documents

The results of examination of the five perspectives, the 3-layer approach, and effect modifiers in the previous section are all finally summarized in Section 2.5.6 Benefits and Risks Conclusions. Therefore, based on the description given in ICH M4 E(r2) Guideline, the items to be noted in each section and their sources in Section 2.5.6 are explained in the subsequent sections.

4.3.1 M2.5.6.1 Background of Treatment

This section should briefly discuss the therapeutic context of the medicinal product. Relevant subgroup differences relevant to the assessment of the biological plausibility of M2.5.6.2 and M2.5.6.3 Layer 2 and Layer 3 should be explained, if known.

ICH M4E(r2)	This section should briefly discuss the therapeutic context for the medicinal product. The term “therapeutic context” describes the disease or condition to be treated, the population intended to be treated, and the benefits and risks of current therapies. Important limitations in the understanding of the condition and uncertainties in the benefits and risks of current therapies should be discussed. If differences in relevant subpopulations are known, they should be discussed. Information on the benefits and risks of the medicinal product should not be included here, but should be discussed in Sections 2.5.6.2 and 2.5.6.3, respectively.

4.3.1.1 M2.5.6.1.1 Disease or Condition

This section provides a description of the aspects of the disease or condition that are most relevant to or have the greatest impact on the intended population. Any differences between regions or relevant subgroups should be described in this section, if known.

ICH M4E(r2)	This section provides a description of aspects of the disease or condition that are most relevant to, or have the greatest impact on, the intended population (e.g., incidence, duration, morbidity, mortality, health-related quality of life). The discussion should focus on the aspects of the disease that would be covered by the proposed indication for the medicinal product. Societal or public health implications of the disease (e.g., impact of poor prevention and control of an infectious disease) should also be addressed where relevant.

4.3.1.2 M2.5.6.1.2 Current Therapies

This section provides a description of the major therapies in the intended population and the medical need for a new therapy in terms of efficacy, safety, tolerability, convenience, or patient preference, if applicable. Drugs in the same pharmacologic class for which there are known differences in relevant subgroups should be described briefly in this section. In addition, any differences in current therapies between regions should be briefly described here.

ICH M4(r2)	This section provides a description of the major therapies in the intended population (i.e., those therapies used most frequently and/or recommended in clinical guidelines) and the medical need for a new therapy in terms of efficacy, safety, tolerability, convenience, or preference, if applicable. For disease areas that are treated by different pharmacologic classes of therapies, this analysis may be simplified by grouping and providing commentary by drug class. Other interventions used for the intended population may also be discussed when their use is supported by established clinical practice or clinical guidelines. Such interventions could include medical and surgical procedures, drugs used off-label, and other non-drug interventions (e.g., diet modifications, physical therapy). Major differences in current therapies between regions may be noted. If no therapies are currently available to treat the intended population, this should be stated.

4.3.2 M2.5.6.2 Benefits

This section provides a factual summary of the data on the key benefits that will be discussed in the benefit-risk assessment of the medicinal product. In order to prepare a CTD based on the principles of ICH E17, this document recommends a systematic summary of the results of studies included in the clinical data package from the five perspectives, using a 3-layer approach, and to evaluate the benefits of the medical product.

The following sections provides the points to be considered in this section using the 3-layer approach and the five perspectives.

ICH M4(r2)	This section provides a factual summary of the data on the key benefits that will be discussed in the benefit-risk assessment of the medicinal product. Benefits are the favorable effects of the medicinal product. In some cases, a benefit may be described by a combination of study endpoints (e.g., the benefit of improved asthma control described by the frequency of exacerbations and hospitalizations and the number of

	asthma-related deaths). If a surrogate endpoint(s) is the basis of the benefit assessment, the ability of the surrogate to predict clinical benefit and the basis for this expectation should be explained. Benefits may also include important characteristics of the medicinal product, such as convenience (e.g., a more convenient dosing regimen or route of administration) that may lead to improved patient compliance, or benefits that affect those other than the patient (e.g., population benefits of a vaccine due to herd immunity). When identifying the key benefits of the medicinal product, the following characteristics should be considered: • Clinical importance of the benefit (e.g., life-prolonging, curative, disease-modifying, symptomatic relief, improved patient compliance, functional or quality of life improvement, prevention of disease progression, prevention of infectious disease, diagnostic). • Magnitude of the absolute difference in frequency of the effect in the study population versus the comparator(s); in some cases, also expressing the difference relative to the comparator may be informative (e.g., if the response rate is 20% in the drug group and 8% in the control group, the absolute difference is 12% (i.e., 20%-8%) and the relative effect is 2.5 (i.e., 20%/8%)). When describing each key benefit, in addition to the points above, the following considerations may also be discussed:. • Time course of the key benefit (e.g., time to onset, continued effect of the product over time). Variability of the key benefit, taking into account relevant subpopulations such as those defined by age, sex, ethnicity, organ function, disease severity, or genetic polymorphism. This section should also include an analysis of the strengths, limitations, and uncertainties of the evidence related to each key benefit and the implications of this information. The following points may be considered, as applicable: • Study design considerations (e.g., superiority or non-inferiority comparison to active control, superiority comparison to placebo, blinding, absence of comparator). • Completeness of data collection and duration of follow-up. • Number of clinical studies and consistency of results across studies. • Relationship between exposure (e.g., drug levels in the blood) and benefit. • Generalizability of the clinical study result to clinical practice (e.g., clinically important differences between the study population and the intended population). • Confidence that surrogate endpoints, if used, predict that the intended population will benefit.

4.3.2.1 Layer 1

In Layer 1, the key benefits of the medicinal product are explained based on the results of the overall population of the studies included in the clinical data package. Layer 1 describes an analysis of the strengths, limitations, and uncertainties of the evidence related to each key benefit and interpretation of this information based on the five perspectives in ICH E17. These five perspectives are also consistent with the elements called for in ICH M4 E(r2).

- Biological Plausibility
  - *Can the results be reasonably explained clinically, pharmacologically, or based on the mechanism of action?*
  - Relationship between exposure (e.g., blood drug concentrations) and the benefit

- Internal Consistency
  - *Are results from biologically or medically relevant endpoints supportive of the same conclusion?*
  - Changes in key benefits over time (e.g., time to response, durability of response)
- External Consistency
  - Number of clinical studies, consistency of results across studies
  - *Consistency with results of similar drugs/similar study populations*

Note: Italicized text indicates additional text considering ICH E17 Guidelines.

- Statistical Uncertainty
  - Discussion of study design (e.g., superiority or non-inferiority to active drug, superiority to placebo, blinded, uncontrolled)
  - Integrity of data collection and follow-up period
  - Reliability of surrogate endpoints to predict benefit in intended patient population
  - *How certain the findings are*
  - *Is there any bias in the estimation that cannot be ignored (considering GCP compliance)*
  - *How high is the precision of estimation*
- Clinical Relevance
  - Clinical significance of benefit (e.g., prolongation of survival, cure, disease modification, symptomatic relief, improvement of patient compliance, improvement of function or quality of life, slowing of disease progression, prevention of infection, diagnosis)
  - Feasibility of clinical study results into routine clinical practice (e.g., clinically important differences between the study population and the patient population for the intended indication)
  - Is the analysis of the strength of evidence, limitations, and uncertainties associated with each primary benefit a meaningful finding that provides the basis for clinical decisions and treatment decisions

4.3.2.2 Layer 2

For Layer 2, in order to appropriately evaluate the strength and limitations of the evidence for the following aspects of ICH M4 E(r2), it is recommended to consider the results of evaluation with effect modifiers as the main axis from the five perspectives for regional and ethnic factors.

- Differences in key benefits considering relevant subgroups such as age, sex, race, organ function, disease severity, and genetic polymorphism

Examples of analysis methods for evaluating effect modifiers from different perspectives and consistency across regions are provided in Section 3.3.1.2. The details of the results of these multifaceted analyses are described in M2.5.4 and M2.7.3.3.3, and in this section, it is recommended to briefly explain the discussion based on the results from the five perspectives using effect modifiers as the main axis.

4.3.2.3 Layer 3

In Layer 3, the key benefits of the medicinal product are evaluated based on the results of Layer 1 and Layer 2 to ultimately evaluate the key benefits in the applicant country/region. If the Layer 2 evaluation does not identify any effect modifiers or candidates, the overall evaluation of the studies can be extrapolated because of the applicant country/region. Once an effect modifier or candidate is identified, it is recommended that the key benefits in the applicant country/region should be evaluated considering the impact of the effect modifier.

The following points should be considered for Layer 3 If an effect modifier or candidate is identified in the results through Layer 2 or if differences are observed between regions.

- If an effect modifier or candidate is identified (Cases B and D in Table 4-1):
  Are there differences or imbalances in intrinsic or extrinsic patient characteristics across countries/regions of application? Also, if there is a difference or imbalance in patient background factors related to the identified effect modifiers or their candidates, will the difference or imbalance affect the benefit in the country/region?
- If there are differences in the population including the country/region of application when evaluating the consistency across regions (Cases C and D in Table 4-1):
  If a difference in treatment effect is observed in the population including the country/region of application for which no effect modifier has been identified or if the difference cannot be explained in terms of other intrinsic or extrinsic ethnic factors, how the difference might affect the benefit in the country/region (what other causes might be considered).
  If a difference in treatment effect is observed in the population including the country/region of application, but the difference can be explained by an effect modifier (identified as an effect modifier) or in terms of other intrinsic or extrinsic ethnic factors, refer to "If an effect modifier or candidate is identified" above.

4.3.3 M2.5.6.3 Risk

This section provides a factual summary of the data on the key risks that will be discussed in the benefit-risk assessment of the medicinal product. In order to prepare a CTD based on the principles of ICH E17, this document recommends a systematic summary of the results of the studies included in the clinical data package from the five perspectives, using the 3-layer approach, evaluating the key risks of the medicinal product, and explaining whether it is possible to monitor, minimize, or control each key risk.

In the following sections, we will summarize the points to be considered in this section using the 3-layer approach and the five perspectives, and present the considerations based on the sources.

ICH M4(r2)	This section provides a factual summary of the data on the key risks that will be discussed in the benefit-risk assessment of the medicinal product. Risks include adverse events and other unfavorable effects associated with the medicinal product. Risks that may be considered also include drug interactions, risks identified in the non-clinical data, risks to those other than the patient (e.g., fetus, those preparing and administering the medicinal product), and risks based on pharmacologic class or current knowledge of the product. Factors such as potential misuse, abuse, or diversion of the product may also be considered. The key risks described in this section may not include all risks that are described elsewhere (e.g., risk management plan, prescribing information). When identifying the key risks of the medicinal product, the following characteristics of risks should be considered: • Seriousness and/or severity • Incidence

	• Reversibility • Tolerability When describing each key risk, in addition to the points above, the following considerations may also be discussed: • Frequencies should generally be presented as the absolute difference relative to the comparator (e.g., placebo, active comparator), and in the context of the background frequency in the patient population. In some cases, also expressing the difference relative to the comparator may be informative. If the frequency is 8% in the treatment arm and 5% in the control group, the absolute difference is 3% (i.e., 8%-5%) and the relative risk is 1.6 (i.e., 8%/5%). • Ability to monitor, minimise, or manage the risk. • Variability of the key risk, taking into account relevant subpopulations such as those defined by age, sex, ethnicity, weight, organ function, disease severity, concomitant illness, concomitant therapy, or genetic polymorphism. • Time course of the adverse event in the study population (i.e., time to onset and resolution, whether the frequency of the event is highest when initiating the drug and subsequently decreases, is relatively constant with time, or increases with cumulative exposure). This section should also include an analysis of the strengths, limitations, and uncertainties of the evidence related to each key risk and the implications of this information. The following points may be considered, as applicable: • Study design considerations (e.g., comparison to active control, comparison to placebo, blinding, absence of comparator). • Adequacy of assessment of risk (e.g., number of patients, number and design of trials, duration of exposure, frequency of monitoring). • Investigation(s) to address safety issues identified during development (e.g., an ophthalmologic investigation conducted to address a non-clinical finding). • Completeness of data collection and duration of follow-up. • Number of patients in relevant subpopulations treated at the intended dose. • Mechanism of action for the adverse event, if known, including non-clinical information or class effects. • Completeness of information on patient characteristics (e.g., smoking history, concomitant medication use) that may affect risk. • Consistency of results across studies. • Relationship between exposure (e.g., drug levels in the blood) and risk. • Generalizability of the clinical study results to clinical practice (e.g., clinically important differences between the study population and the intended population). The proposed approach to managing each key risk should also be discussed, including an explanation of why the approach provides reasonable assurance that the risk can be appropriately managed. Repetition of details from the risk management plan is not necessary. In certain cases, a discussion of the overall approach to risk management may be sufficient and may be included after all key risks have been identified and described.

4.3.3.1 Layer 1

For Layer 1, the key risks of the medicinal product are described based on the results of the overall population of the studies included in the clinical data package. Layer 1 describes the strength of evidence associated with each key risk, an analysis of limitations and uncertainties,

and interpretation of this information based on the five perspectives in ICH E17. These five perspectives are also consistent with the contents called for in ICH M4 E(r2).

- Biological Plausibility
  - *Can the results be reasonably explained clinically, pharmacologically, or based on the mechanism of action*
  - Known mechanisms of adverse events, including nonclinical information or information on effects common to drugs in the same therapeutic class
  - Relationship between exposure (e.g., blood drug concentrations) and risk
- Internal Consistency
  - *Consistent results on relevant endpoints (e.g., adverse events and laboratory values)*
  - Changes in adverse events over time, reversibility
- External Consistency
  - Consistency of results among studies
  - *Consistency with results of similar drugs/similar study populations*
  - *Consistency with results in other indications of the medicinal product*
  - Review to address safety issues identified during drug development (e.g., ophthalmologic examinations designed to address nonclinical findings)
- Statistical Uncertainty
  - Discussion of study design (e.g., comparison to active, comparison to placebo, blinded, uncontrolled)
  - Appropriateness of risk assessment (e.g., number of patients, number of studies and study design, duration of exposure, frequency of monitoring)
  - Integrity of data collection and follow-up period
  - Number of patients in relevant subgroups who received the proposed dosage and administration
  - Information of integrity about patient characteristics affecting risk (e.g., smoking history, concomitant therapy)
- Clinical Relevance
  - Feasibility of clinical study results into routine clinical practice (e.g., clinically important differences between the study population and the patient population for the intended indication)
  - Is the analysis of the strength, limitations, and uncertainties of the evidence associated with each key risk a key finding that provides the basis for clinical decisions and treatment decisions?

4.3.3.2 Layer 2

For Layer 2, in order to appropriately evaluate the strength and limitations of the evidence for the following aspects of ICH M4 E(r2), it is recommended to consider the results of the evaluation with effect modifiers as the main axis from the five perspectives for regional and ethnic factors.

- Differences in key risk considering relevant subgroups such as age, sex, race, weight, organ function, disease severity, co-morbidities, concomitant therapy, or genetic polymorphism

Examples of analysis methods for evaluating effect modifiers from different perspectives and consistency across regions are provided in Section 3.3.2. The details of the results of these multifaceted analyses are described in M2.5.5, M2.7.4.5.1, and M2.7.4.5.2. In this section, it is recommended to briefly explain the discussion based on the results from the five perspectives using effect modifiers as the main axis.

4.3.3.3 Layer 3

In Layer 3, the key risks of the medicinal product are evaluated based on the results of the Layer 1 and Layer 2 in order to ultimately evaluate the key benefits in the applicant country/region. If the Layer 2 evaluation does not identify an effect modifier for safety or a candidate safety effect modifier, the overall study evaluation of the studies can be extrapolated because of the applicant country/region. Once a safety effect modifier or candidate is identified, it is recommended that key risks in the applicant country/region should be evaluated considering the impact of the effect modifier.

If effect modifiers or their candidates are identified in the results up to Layer 2, or differences are observed across regions, points to be considered for Layer 3 are shown in Section 4.3.2.3.

4.3.4 M2.5.6.4 Benefit-Risk Assessment

In this section, the benefit-risk assessment discussed in Section 4.3.2 and Section 4.3.3 is briefly concluded considering the balance for each layer. In addition to describing the evaluation of each layer individually, we recommend that the composition should be consistent throughout the evaluation results from Layer 1 to Layer 3. Others are described in accordance with ICH M4 E(r2).

5 Example

This section provides the description examples of the examination and discussion of Layer 2 and Layer 3 for efficacy and benefit in Module 2.5 or Module 2.7.3 of the CTD as a reference for a benefit-risk explanation based on multifaceted and systematic consistent evaluation of efficacy and safety across populations or regions. The description example is based on ticagrelor (Section 5.1) and 2 virtual examples (Section 5.2 and Section 5.3) presented in ICH E17 Training Material Module 6, which show background information, a description example of CTD, and a summary from the five perspectives, respectively. The background information and study results of these 3 cases were based on review-related information and published materials [Carroll and Fleming 2013, 7th ICH E17 Workshop (2022)], but they were fictitious settings that were simplified to facilitate sharing of images described in the CTD, and the accuracy, appropriateness, and reality related to the target disease were not considered in virtual examples.

5.1 Example 1: Ticagrelor

5.1.1 Background Information

Indication	Acute coronary syndrome
Item	Ticagrelor
Country of application	United States
MRCT in the submission package	1 study Multi-regional phase III study PLATO study (Pivotal study)
Pivotal Study Design	
Design Overview	A randomized, double-blind, clopidogrel-controlled, multi-regional phase III study Participating countries: United States and 43 other countries in total
Treatment arm	Ticagrelor (Hereinafter referred to as this drug group), clopidogrel (Hereinafter referred to as control arm) Randomization ratio 1:1, Stratification factors: None In both arms, an aspirin maintenance dose 75-100 mg/day was concomitantly administered as the basal treatment (in patients with stent placement, 325 mg/day was allowed until 6 months after the implantation).
Treatment duration	6-12 months
Primary Endpoint	Time to first occurrence of any of the composite events of cardiovascular death, myocardial infarction, and stroke
Secondary Endpoints	Time to first occurrence of each component of the primary endpoint Safety Pharmacokinetics (PK)
Number of Participants	18624 subjects (approximately 9300 subjects per treatment arm), including 1413 subjects from the United States (approximately 700 subjects per treatment arm)
Efficacy Results Across Pivotal Studies (Layer 1)	• The primary endpoint of incidence of composite events was significantly lower in the test drug arm than in the control arm (hazard ratio 0.84). • The incidences of cardiovascular death and myocardial infarction were significantly lower in the test drug arm than in the control arm, and the incidence of stroke did not differ significantly between the arms.
Other	• The test drug (ticagrelor) is a similar drug to the control drug clopidogrel. • There are no known clinical or pharmacological features from early clinical studies or published literature that would predict different efficacy of the test drug in a particular population. • The recommended maintenance dose range of aspirin is different between the US clinical practice guidelines (75-325 mg/day) and the non-US clinical practice guidelines (75-100 mg/day) at the time of planning the study. • In clinical practice, there may be no regional differences in the treatment or clinical course of acute coronary syndrome.

5.1.1 Description Example in CTD

5.1.1.1 Description example in M2.5.4 or M2.7.3.3.3

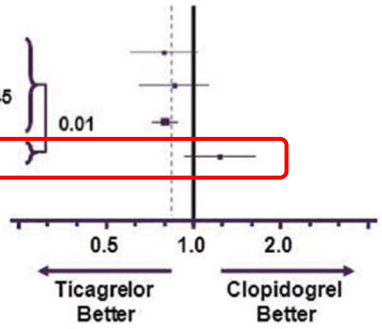
M2.X.X.X Efficacy in Special Populations

M2.X.X.X.1 Subgroup Analysis of the Primary Endpoint

Subgroup analyses of the primary endpoint were planned and performed for 31 background factors (intrinsic, extrinsic, and geographic region). As a result, for background factors other than geographical region, there was no tendency toward obvious difference of the efficacy of the drug between the subgroup (Appendix-Figure X-X_ not shown). When stratified by geographical region, the point estimate of the hazard ratio exceeded 1 in North America, while it was below 1 in all 3 regions except North America (US, Canada), showing a difference in efficacy between North America and non-North America (Figure example 1-1).

Figure example 1-1 Primary Endpoint by Region – Forest Plot (PLATO study, FAS)

Characteristic	Total Patients	KM at Month 12 Tic	KM at Month 12 Clop	HR (95% CI)	Interaction p-values
Geographic Region					
Asia / Australia	1714	11.4	14.8	0.80 (0.61, 1.04)	0.045
Cent / Sth America	1237	15.2	17.9	0.86 (0.65, 1.13)	
Euro / Md E / Afr	13859	8.8	11.0	0.80 (0.72, 0.90)	0.01
North America	1814	11.9	9.6	1.25 (0.93, 1.67)	



Source: Carroll and Fleming 2013
HR = hazard ratio; KM = Kaplan-Meier.

M2.X.X.X.2 Assessment of the Impact of Primary Endpoint for Participating Countries by Post-hoc Additional Analyses

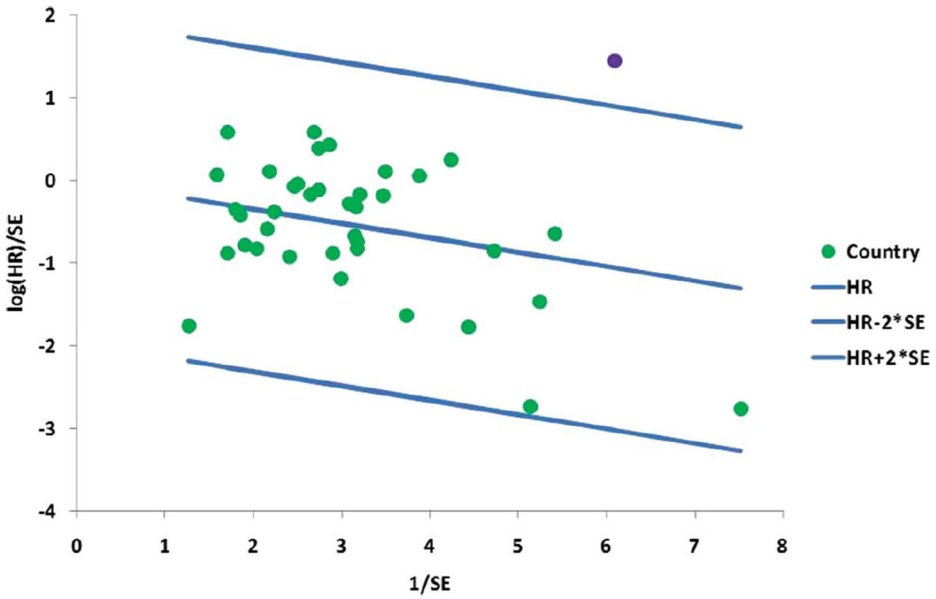
To explore factors for which differences were observed among geographical regions, the following post-hoc additional analyses were performed for the primary endpoint. Upon examining the matters related to study operation (e.g., study procedures, transportation of investigational product, site management status, dropout rate, protocol deviation rate) in detail, it was considered that these were not the causes of differences in efficacy among the geographical regions.

(1) Assessment by participating country

Distribution was visually assessed by plotting the results of the primary endpoint for each country. As a result, in both the Galbraith plot and the Q-Q plot, only the US (hazard ratio of 1.27) diverged from the distribution, showing heterogeneity of the results in the US (Figure example 1-2, Figure example 1-3). In the funnel plot, only the US diverged from the pseudo 95% CI (Figure example 1-4), suggesting that there was some systematic factor for the

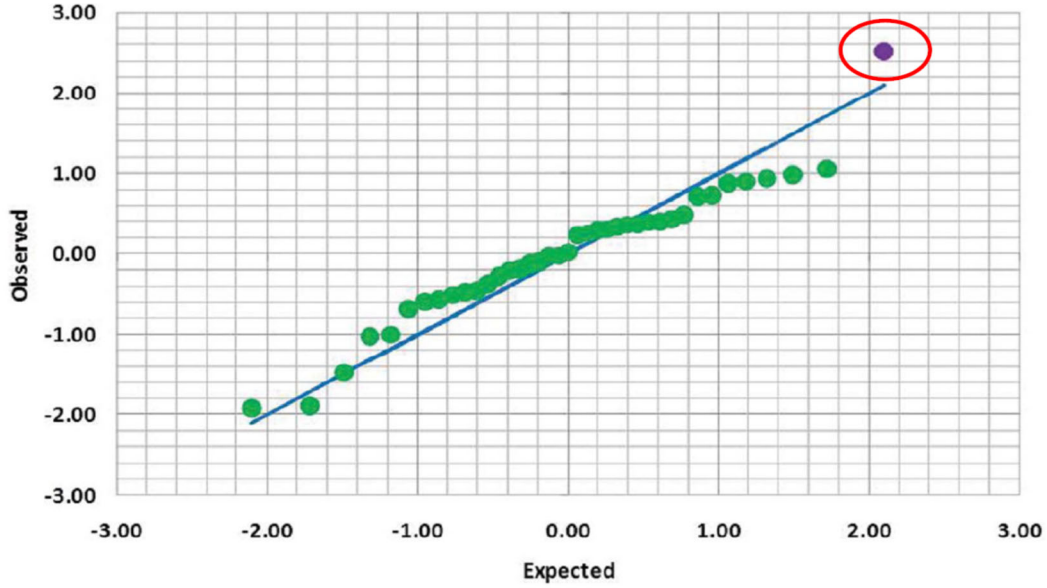
result in the US. The US had the second largest number of enrollment participants and events in this study (151 of 1878 events), the treatment-by- the US interaction was significant ($p = 0.0095$). Also, the probability that the hazard ratio of the primary endpoint would become ≥ 1.27 in the US when it was 0.84 in the overall population was < 0.006 (Appendix-Table X-X_ not shown). However, the study is not powered to verify the results in the US alone, and the possibility that the results in the US are incidental cannot be ruled out.

Figure example 1-2 Primary endpoint by country–- Galbraith Plot (PLATO study, FAS)



Source: Carroll and Fleming 2013
SE = standard error. Red circle indicates US.

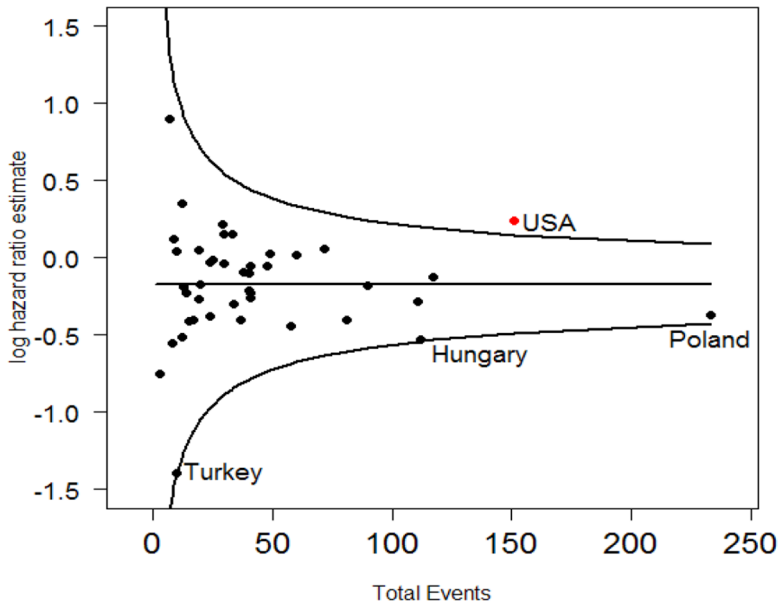
Figure example 1-3 Primary endpoint by country–- Q-Q Plot (PLATO study, FAS)



Source: Carroll and Fleming 2013

Horizontal axis: normal distribution, vertical axis: observed values. Red circle indicates US.

Figure example 1-4 Primary endpoint by country–- funnel plot (PLATO study, FAS)



Source: Ticagrelor US Review Related Information (2010)

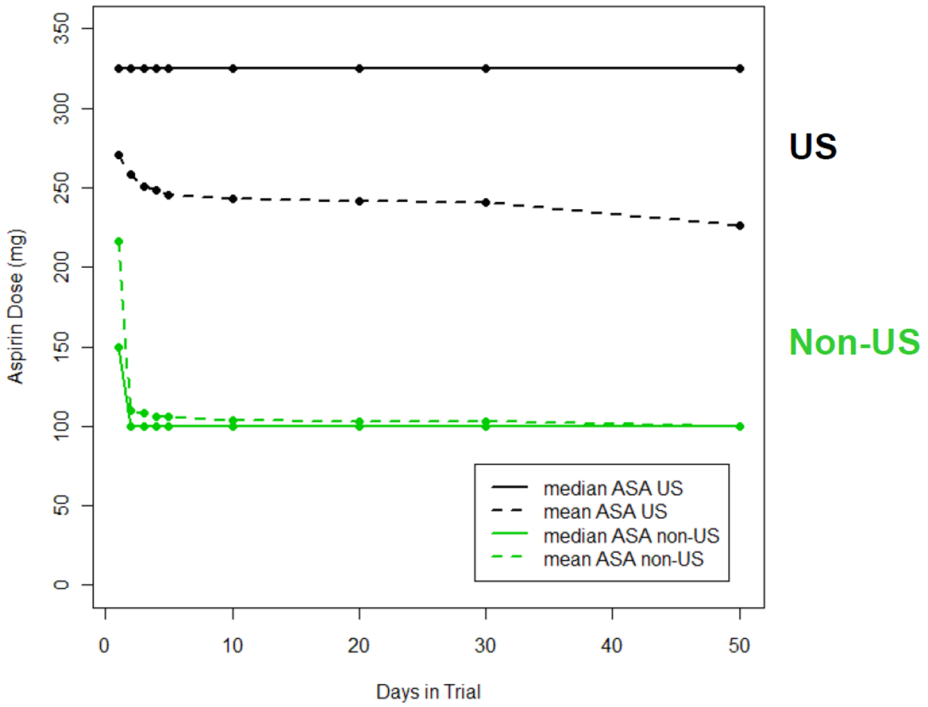
(2) Explore factors contributing to the differences in the efficacy results of the drug between the US and non-US

In order to explore factors contributing to the differences in the efficacy results of the test drug between the US and non-US, imbalances of characteristics in baseline and after baseline between the regions were examined. After performing Cox regression analysis with treatment arm, region (US, non-US), and interaction between treatment arm and region as covariates, the interaction between treatment arm and each characteristic was added to the model and analyzed. These results showed that a clear interaction in treatment arm and the characteristic with disproportionality between the US and non-US populations was the dose of concomitant aspirin as basic care (treatment arm by aspirin dose interaction: $p = 0.003$). Mean and median aspirin doses were higher in the US than in non-US subjects throughout the study (Figure example 1-5), with the percentage of subjects with doses > 100 mg being 57% in the US and 8% in non-US subjects, and doses ≥ 300 mg being 54% in the US and 2% in non-US subjects. The proportions of subjects by aspirin dose were similar between the treatment arms in the US and non-US (Appendix-Table X-X_ not shown).

Subgroup analyses of the primary endpoint by aspirin dose in the US and non-US showed similar results in the US and non-US. In the US, the number of participants in the mid-dose group was small and the hazard ratio could not be estimated, but the hazard ratio was higher in the high-dose group than in the low-dose group, and the higher the dose, the higher the hazard ratio in non-US. In both the US and non-US, the point estimate of the hazard ratio was below 1 only in the group of ≤ 100 mg (Figure example 1-6). Therefore, the dose of aspirin as the basic care is an effect modifier of the drug, and the reason why the efficacy of the test drug was not demonstrated only in the US seems to be the high proportion of the participants who

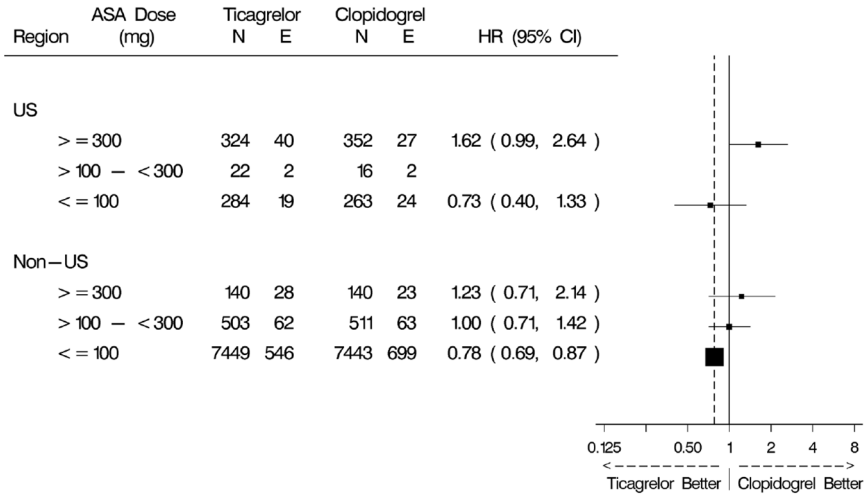
concomitantly used a high dose of aspirin in the US. However, these are post-hoc additional analyses and there are limitations that multiplicity is not adjusted and the sensitivity of the model is high (influenced by the number of participants in non-US who received high doses of aspirin and the handling of missing aspirin data).

Figure example 1-5 Changes in aspirin dose in the US and non-US (PLATO study, FAS)



Source: Ticagrelor US Review Related Information (2010)

Figure example 1-6 Primary endpoint by aspirin dose in US and non-US – forest plot (PLATO study, FAS)



Source: Ticagrelor US Review Related Information (2010)
E = events; N = number of patients.

The mechanism of action also theoretically suggests that concomitant use of higher doses of aspirin may result in reduced efficacy. At low doses, aspirin inhibits COX-1 and inhibits platelet aggregation, but at high doses it also inhibits endothelial COX-2, resulting in increased vascular resistance. In addition, the test drug almost completely blocks $P2Y_{12}$ receptors and appears to inhibit platelet aggregation mediated by COX-1 inhibition, limiting the antiplatelet effect of aspirin (Warner et al. 2011).

5.1.1.2 Description example in M2.5.6.2

M2.5.6.2 Benefits

< Describe after the benefit claim based on overall study results >

Based on the results of early clinical studies and the PPK analysis in the PLATO study, no intrinsic or extrinsic ethnic factors were found to have a clinically significant effect on the PK of the test drug in patients with acute coronary syndrome. In addition, the results of subgroup analysis of the primary endpoint in the PLATO study did not suggest any tendency for the efficacy of the test drug to clearly differ depending on the participant background factors except for geographic region.

The primary endpoint by geographic region in the PLATO study was that the point estimate of the hazard ratio was less than 1 in Asia/Australia, South America/Central America, and Europe/the Middle East/Africa, while it was greater than 1 in North America (the US and Canada). As a result of examining the influence of participating countries on the efficacy by multiple post-hoc additional analyses, there was heterogeneity in the results from the US, and it was suggested that there was some systematic factor for the results from the US. To investigate this factor, we explored the characteristics that clearly interact with the treatment arm and are imbalanced between the US and non-US and found that the dose of aspirin used as the basic care corresponded.

Subgroup analyses of the primary endpoint by aspirin dose (daily dose ≤ 100 mg, > 100 to < 300 mg, ≥ 300 mg) were performed in the US and non-US. The results in the US and non-US were similar, suggesting that the hazard ratio tended to be higher with higher aspirin dose. The point estimate of the hazard ratio was below 1 only at the dose of 100 mg or less in both in the US and non-US. Although no clear benefit of the test drug was shown at the dose of > 100 mg, a consistent benefit was suggested at the dose of 100 mg or less in both in US and non-US. Therefore, the dose of aspirin as the basic care is an effect modifier of the drug, and the reason why the efficacy of the test drug was not demonstrated only in the US seems to be the high proportion of the participants who concomitantly used a high dose of aspirin in the US. Although there is a limitation that multiplicity is not adjusted and the sensitivity of the model is high since these are post-hoc additional analyses, the benefit of the test drug can be expected to consistently exceed that of the control drug in each country when used concomitantly with low-dose aspirin (75-100 mg/day).

In the multi-regional including North America phase III PEGASUS study in patients with old myocardial infarction, no regional difference was observed in the efficacy of the test drug under concomitant use with low-dose aspirin (75-100 mg/day), supporting the consistency in the efficacy of the test drug between the regions (Bonaca et al. 2015).

The efficacy of the test drug in the US was found to be inferior to that in the non-US in the PLATO study. However, concomitant use with low-dose aspirin (75-100 mg/day) as mentioned above is expected to provide benefits over the control drug in US patients.

5.1.2 Summary of the Five Perspectives

Biological Plausibility	The mechanism of action also suggests that high-dose aspirin may affect the test drug efficacy. There are no intrinsic ethnic factors that clinically influence the PK of test drug.
Internal Consistency	The efficacy of the test drug by aspirin dose was similar in the US and non-US.
External Consistency	In studies of similar diseases, the test drug was consistently effective when used concomitantly with low-dose aspirin across regions.
Statistical Uncertainty	Although the possibility of a chance result cannot be ruled out, the heterogeneity of the results in the US was supported by multiple assessment methods. Although the interaction of aspirin dose was significant, it has a limitation of being a post-hoc additional analysis, multiplicity-unadjusted, and high sensitivity model.
Clinical Relevance	The risk of cardiovascular events is lower than that of the control drug only if the concomitant aspirin dose is low.

5.2 Example 2: Virtual Example ABC 123

5.2.1 Background Information

Indication	Breast cancer
Item	ABC123
Country of application	Japan
MRCT in the submission package	1 study Multi-regional phase III study 301 (Pivotal study)
Pivotal Study Design	
Design Overview	A randomized, double-blind, DEF456 controlled, multi-regional phase III study Participating countries: 6 countries (China, US, Germany, Japan, France, Italy)
Treatment arm	ABC123 (Hereinafter referred to as test drug arm), DEF456 (hereinafter referred to as control arm) Randomization ratio 1:1, stratification factors: lesion site (visceral metastasis present, no visceral metastasis), history of adjuvant surgery (with or without prior hormone therapy)
Treatment duration	Not fixed (until progression or other discontinuation criteria are met) *After treatment discontinuation, patients will be followed for survival until death or another discontinuation criterion is met.
Primary Endpoint	Progression-free survival (PFS)
Secondary Endpoints	Overall survival (OS), overall response rate (ORR), duration of response (DOR) Safety

	PK
Number of Participants	600 subjects (300 subjects per arm), including 60 Japanese subjects (30 subjects per arm)
Efficacy Results Across Pivotal Studies (Layer 1)	• OS was significantly prolonged in the test drug arm compared with the control arm. • PFS, ORR, and DOR were all better in the test drug arm than in the control arm.
Other	• The test drug (ABC123) is a similar drug to the control drug DEF456. • There seems to be no regional difference in the medical environment such as diagnosis and treatment of breast cancer. • There are no known clinical or pharmacological features from early clinical studies or published literature that would predict different efficacy of the test drug in a particular population.

5.2.2 Description Example in CTD

5.2.2.1 Description example in M2.5.4 or M2.7.3.3.3

M2.X.X.X Efficacy in special patient populations

M2.X.X.X.1 Subgroup Analyses of PFS and OS

PFS and OS were examined by age, lesion site, and history of hormone therapy, which are prognostic factors for breast cancer, in addition to analysis by race. The results showed no tendency toward difference in the efficacy of the test drug among subgroups. Although the 95% CI was wide for races other than Asian or Caucasian, in which the number of participants was limited to 36, the point estimate of the hazard ratio versus the control arm was below 1 for both PFS and OS in all subgroups, which was consistent with the overall result (Figure example 2-1, Figure example 2-2).

Figure example 2-1 PFS by participant background factor (Study 301, ITT population)

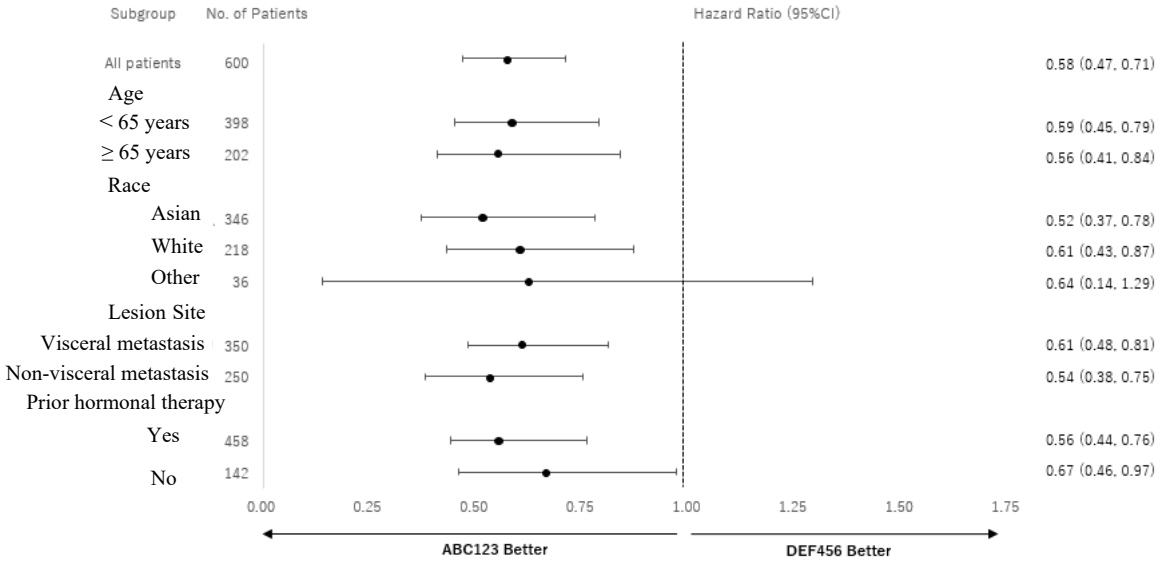
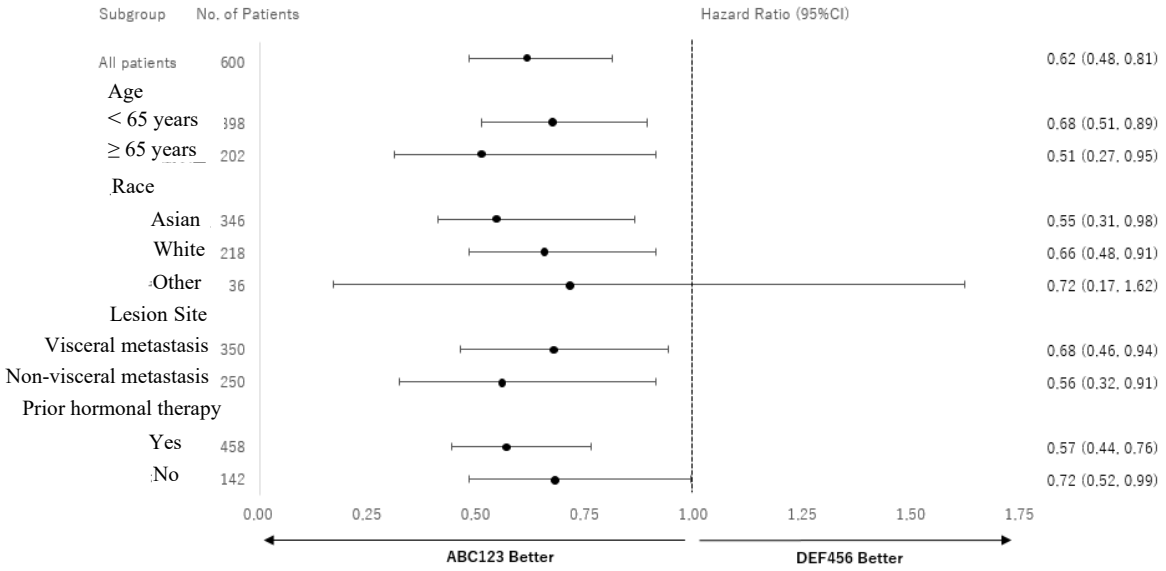


Figure example 2-2 OS by participant background factor (Study 301, ITT population)

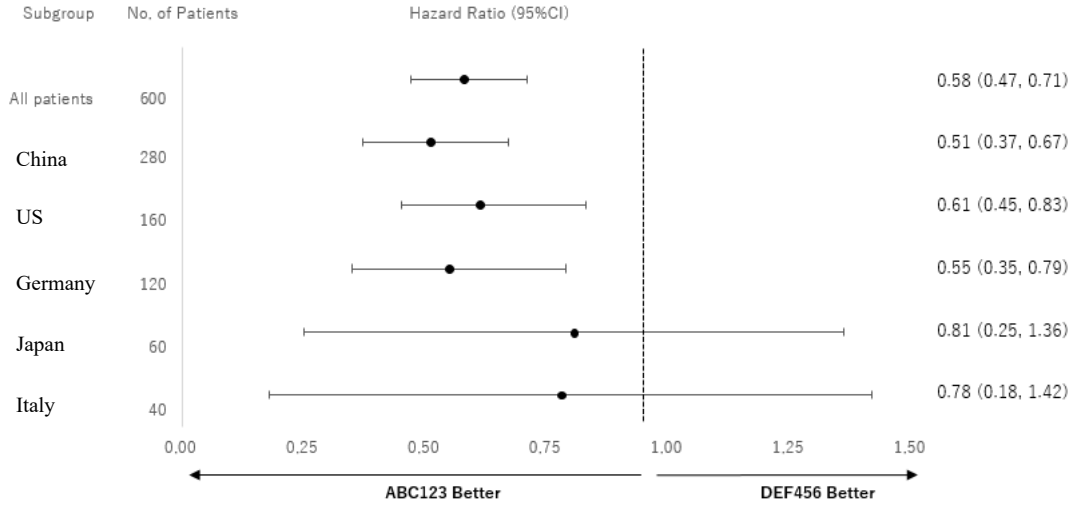


M2.X.X.X.2 Impact on PFS in participating countries

To evaluate the consistency of the efficacy of the test drug among participating countries, analysis of PFS by participating country, the Gail-Simon test, and analysis using the J-S shrinkage estimator were conducted.

The results of analysis of PFS by participating country showed no tendency of difference in the efficacy of the test drug among the participating countries. Although the 95% CI was wide in Japan and Italy where the number of participants was limited, the point estimate of the hazard ratio versus the control arm was below 1 in all countries, which was consistent with the overall result (Figure example 2-3).

Figure example 2-3 PFS by participating country – Forest plot (Study 301, ITT population)



When the Gail-Simon test was used to test the efficacy in the test drug arm as compared with the control arm and the interaction in the participating countries, the result suggested that there is no qualitative interaction between the efficacy and the participating countries ($p = 0.777$), and the consistency of efficacy among the participating countries was not ruled out (Appendix-Table X-X not shown).

The weighted mean (shrinkage estimate) of the estimate calculated using a J-S shrinkage estimator, based on the efficacy of the overall test drug arm as compared with the control arm and data of each countries showed a similar trend as overall result in all countries, and consistency of efficacy among the countries participating in the study of this study was not ruled out (Appendix-Figure X-X not shown).

5.2.2.2 Description example in M2.5.6.2

M2.5.6.2 Benefits

< Describe after the benefit claim based on overall study results >

Based on the results of PPK analysis of early clinical studies in Japan and non-Japan and a multi-regional phase III study (Study 301), no intrinsic or extrinsic ethnic factors that have a clinically significant effect on PK were identified. In Study 301, PFS and OS were evaluated by age, lesion site (presence or absence of visceral metastasis), and history of hormone therapy, which are prognostic factors for breast cancer, in addition to race. The results showed no tendency toward difference in efficacy of the test drug between subgroups, although the number of participants was small and variability was large in some subgroups. Therefore, the test drug is expected to provide consistent benefits regardless of these background factors.

Furthermore, the effect of participating countries on the efficacy of the test drug in PFS was examined. Although the number of participants was small in some countries and the variation was large, the consistency of the efficacy among participating countries was not ruled out in

any of analyses. Therefore, the results of the overall population are applicable to each participating country, and the test drug is expected to provide consistent benefits to each country.

Although PFS in Japanese participants had a wide 95% CI, the point estimate of the hazard ratio was below 1. As described above, since the consistency of the efficacy of the test drug among the participating countries could not be ruled out, the benefit of the test drug is expected to exceed that of the control drug also in Japanese patients.

The efficacy of DEF456, a similar drug, for breast cancer is not known to differ between Japan and overseas.

5.2.3 Summary of the Five Perspectives

Biological Plausibility	The efficacy of DEF456, a similar drug, for breast cancer is not known to differ between Japan and overseas. There are no known clinical or pharmacological features from early clinical studies or published literature that would predict different efficacy of the test drug in a particular population.
Internal Consistency	In Study 301, efficacy of the test drug did not tend to differ between subgroups in terms of either OS or PFS when analyzed by patient characteristics.
External Consistency	The efficacy of DEF456, a similar drug, for breast cancer is not known to differ between Japan and overseas. There are no known clinical or pharmacological features from early clinical studies or published literature that would predict different efficacy of the test drug in a particular population.
Statistical Uncertainty	When PFS was examined by participant background factors in Study 301, the number of participants was small in some subgroups, showing large variability. In the analysis of PFS by country in Study 301, the number of participants was small in some countries including Japan, showing large variability.
Clinical Relevance	It is considered that there are no regional differences in the medical environment such as diagnosis and treatment of breast cancer. And the consistency of the efficacy of the test drug among the participating countries was not denied in the evaluation of Study 301.

5.3 Example 3: Virtual Example GHI789

5.3.1 Background Information

Indication	Non-radiographic axial spondyloarthritis
Item	GHI789
Country of application	Japan
MRCT in the submission package	1 study Multi-regional phase III study 301 (Pivotal study)
Pivotal Study Design	
Design Overview	A randomized, double-blind, placebo-controlled, multi-regional phase III study Participating countries: 15 countries (US, Bulgaria, Spain, Denmark, Germany, Belgium, UK, Switzerland, China, Netherlands, France, Japan, Sweden, Brazil, Canada)

Treatment arm	GHI789 (Hereinafter referred to as test drug arm), Placebo (Hereinafter referred to as control arm) Randomization ratio 1:1, stratification factors: C-reactive protein (CRP) level or presence or absence of inflammatory signs by magnetic resonance imaging (MRI) [A CRP level above the normal range with sacroiliitis on MRI (CRP+/MRI+), a CRP level above the normal range without sacroiliitis on MRI (CRP+/MRI-), a CRP level within the normal range with sacroiliitis on MRI (CRP-/MRI+)] * *This study was not indicated for patients with normal CRP levels and no sacroiliitis on MRI (CRP-/MRI-).
Treatment duration	Week 52
Primary Endpoint	ASAS40 response rate * at Week 16 *Proportion of participants with improvement in clinical symptoms/signs, as defined by results from multiple scale assessments
Secondary Endpoints	Improvement evaluation of clinical symptoms/signs such as ASAS20 response rate, improvement evaluation of physical function Safety PK
Number of Participants	350 subjects (175 subjects per arm), including 8 Japanese subjects (2 subjects in the test drug arm and 6 subjects in the control arm)
Efficacy Results Across Pivotal Studies (Layer 1)	• The ASAS40 response rate at Week 16 was significantly higher in the test drug arm than in the control arm. • The secondary endpoints also showed favorable efficacy in the test drug arm compared to the control arm.
Other	• Published literature suggests that men with axial spondyloarthritis are at increased risk of disease progression and are more responsive to drugs. There are no known racial or regional differences in the ratio of male to female patients. • Published literature suggests that patients with axial spondyloarthritis who have elevated CRP or inflammatory signs on MRI are at increased risk of disease progression and are more responsive to drugs. There are no known racial or regional differences in the proportion of patients with inflammatory signs. • There seems to be no regional difference in the medical environment such as diagnosis and treatment of axial spondyloarthritis.

5.3.2 Description Example in CTD

5.3.2.1 Description example in M2.5.4 or M2.7.3.3.3

M2.X.X.X Efficacy in special patient populations

M2.X.X.X.1 Patient demographics and analysis by country

Factors that may affect the efficacy of the test drug in patients with axial spondyloarthritis were sex, CRP level, and presence or absence of inflammation on MRI. By sex, it has been suggested that male patients are at higher risk of disease progression and are more responsive to drugs (reference). It has also been suggested that patients with high CRP levels or inflammatory signs based on MRI findings are also at high risk of disease progression and show high drug response

(reference). There are no known racial or regional differences in these patient background factors.

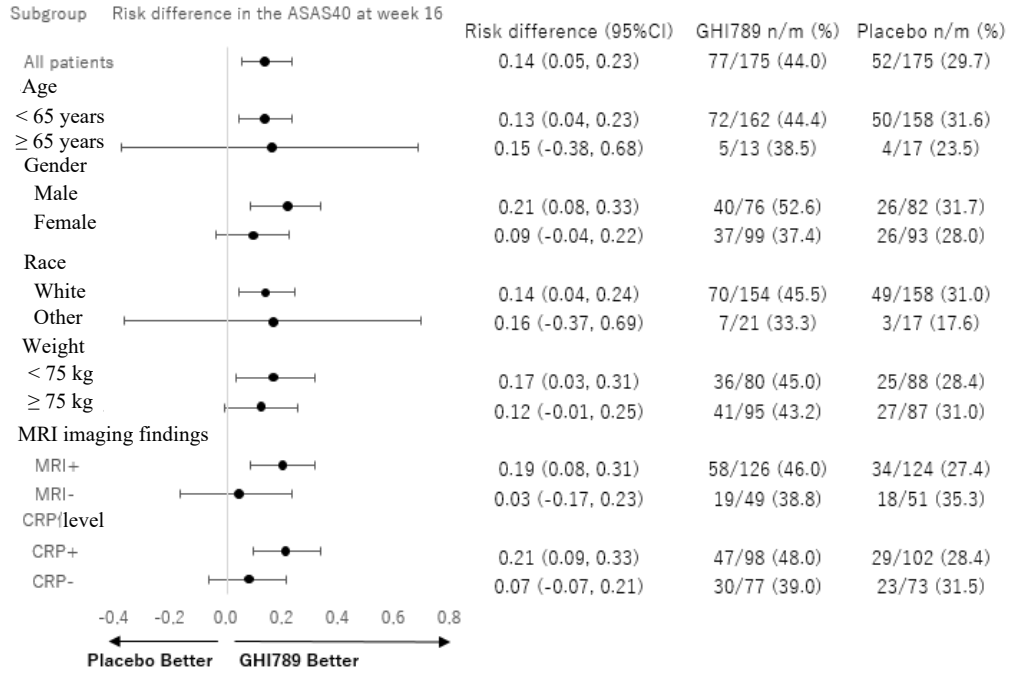
The ASAS40 response rate at Week 16 was examined by participant background factors, including these factors and by country of this study participation.

- **Analysis by participant background factor**

The ASAS40 response rate at Week 16 was higher in the test drug arm than in the control arm for all subgroups stratified by age, sex, race, body weight, CRP level, and MRI findings.

The between-arm difference in the ASAS40 response rate at Week 16 between the test drug arm and the control arm was greater in males than females, CRP+ than CRP-, and MRI+ than MRI-, suggesting that sex, CRP level, and the presence or absence of inflammatory signs on MRI images may be effect modifiers of the test drug. There was no significant difference between the treatment arms in terms of body weight, suggesting no influence of body weight on the efficacy of the test drug. The subjects were younger than 65 years (91.4%), and most of them were white (89.1%), precluding accurate evaluation by age and race, but there was no tendency for the efficacy of the test drug to clearly differ (Figure example 3-1).

Figure example 3-1 ASAS response rate at Week 16 by participant background factor (Study 301, ITT population)



n: number of responders, m: number of patients

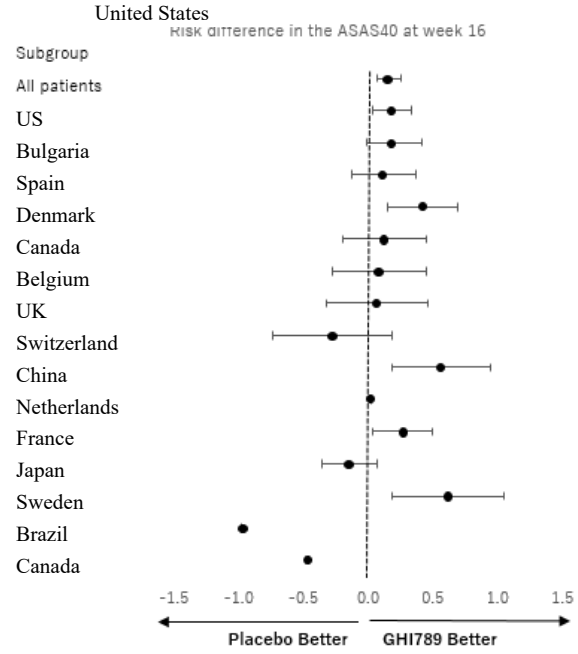
- **Review by participating country**

Due to the limited number of participants (less than 20) in 9 of the 15 participating countries, it was difficult to assess the consistency among the participating countries. The difference in the

ASAS40 response rate at Week 16 between the test drug and control was ≤ 0 in 5 of 15 countries, all of which had < 20 participants (Figure example 3-2).

For the ASAS40 response rate at Week 16, there was no significant difference in the test of interaction between participating countries ($p = 0.3210$), and consistency of efficacy among participating countries could not be ruled out (Appendix-Table X-X not shown).

Figure example 3-2 ASAS response rate at Week 16 by participating countries (Study 301, ITT population)



Subgroup	Risk difference (95%CI)	GHI789 n/m (%)	Placebo n/m (%)
All patients	0.14 (0.05, 0.23)	77/175 (44.0)	52/175 (29.7)
US	0.16 (0.01, 0.31)	18/28 (64.3)	11/23 (47.8)
Bulgaria	0.18 (-0.03, 0.39)	11/19 (57.9)	10/25 (40.0)
Spain	0.10 (-0.15, 0.35)	12/22 (54.5)	9/20 (45.0)
Denmark	0.40 (0.13, 0.67)	12/20 (60.0)	4/20 (20.0)
Canada	0.11 (-0.21, 0.43)	4/16 (25.0)	4/22 (20.0)
Belgium	0.07 (-0.29, 0.43)	3/14 (21.4)	2/14 (14.3)
UK	0.05 (-0.34, 0.44)	4/12 (33.3)	2/7 (28.6)
Switzerland	-0.30 (-0.76, 0.16)	2/10 (20.0)	4/8 (50.0)
China	0.55 (0.17, 0.93)	6/11 (54.5)	0/5 (0)
Netherlands	0.00 (0.00, 0.00)	0/5 (0)	0/11 (0)
France	0.25 (0.02, 0.48)	2/4 (50.0)	1/8 (12.5)
Japan	-0.17 (-0.38, 0.05)	0/2 (0)	1/6 (16.7)
Sweden	0.60 (0.17, 1.03)	3/5 (60.0)	0/2 (0)
Brazil	-1.00 (-1.00, -1.00)	0/5 (0)	2/2 (100)
Canada	-0.50 (-0.50, -0.50)	0/2 (0)	1/2 (50.0)

n: number of responders, m: number of patients

M2.X.X.X.2 Evaluation by group based on effect modifiers by post-hoc additional analysis

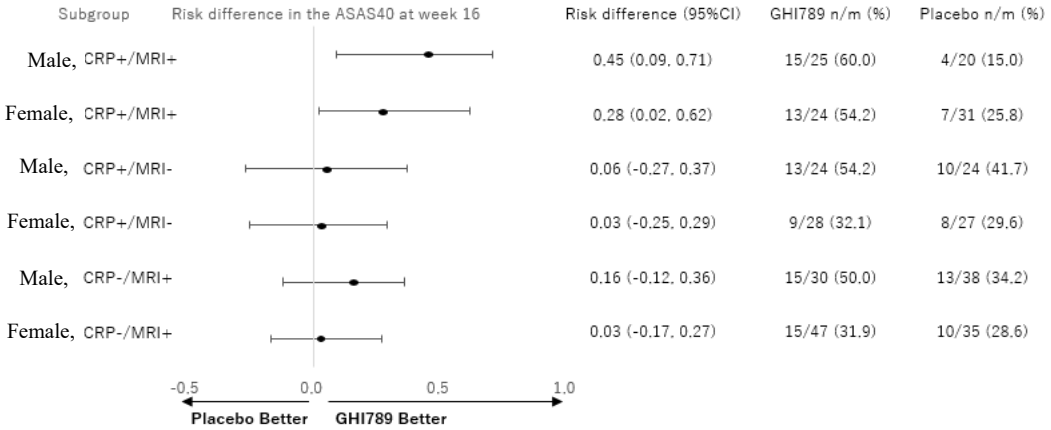
In the review described in the previous section, sex and the presence or absence of inflammatory signs on CRP and MRI findings were considered as candidate effect modifiers. Also, the number of participants was limited in multiple participating countries, which made it difficult to evaluate the consistency of efficacy among the participating countries. To further evaluate the consistency of efficacy, a post hoc additional analysis was performed to examine the ASAS40 response rate at Week 16 in similar participants or countries grouped by potential effect modifiers.

- **Assessment by Participant Group**

When participants were grouped into 6 groups by sex (male, female) and CRP level or inflammatory signs based on MRI findings (CRP+/MRI+, CRP+/MRI-, CRP-/MRI+), the ASAS response rate at Week 16 was higher in the test drug arm than in the control arm across all groups.

The between-arm differences in each group suggested that sex and the presence or absence of inflammatory signs on CRP or MRI were effect modifiers of the test drug, particularly highly influential by the presence of inflammatory signs on both or only one of CRP or MRI. In both sexes, the between-arm difference was greater in the CRP+/MRI+ groups than in the CRP+/MRI- groups or the CRP-/MRI+ groups. The difference of the between-arm difference between male and female was small, but greater in men in the CRP+/MRI+, the CRP+/MRI-, and the CRP-/MRI+ groups.

Figure example 3-3 ASAS response rate at Week 16 by participant group (Study 301, ITT population)



n: number of responders, m: number of patients

- **Assessment by group of participating countries**

The participating countries were grouped into 4 groups in consideration of the proportion of CRP+/MRI+ and male participants in each participating country and the balance of the number of participants. The proportion of CRP+/MRI+ and male participants by group of participating countries is shown in Table example 3-1, and the ASAS response rate at Week 16 is shown in Figure example 3-4.

Groups 1 and 2 had a higher proportion of CRP+/MRI+ participants than overall, and Groups 1 and 3 had a higher proportion of male participants than overall. Among the groups, both CRP+/MRI+ and the proportion of male participants were highest in Group 1 and lowest in Group 4. In all groups, the proportion of participants with CRP+/MRI+ and male participants were similar between the two arms (Appendix-Table X-X not shown).

For all groups of participating countries, the ASAS response rate at Week 16 was higher in the test drug arm than in the control arm.

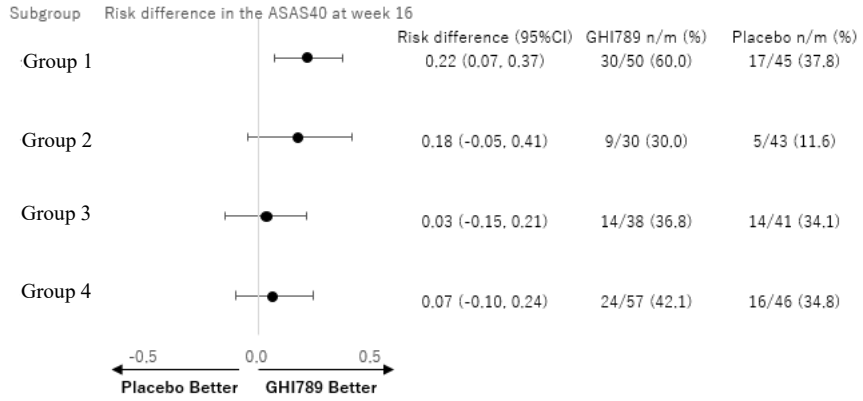
Since the between-arm differences between groups of participating country were similar, the results of the overall population were considered applicable to each participating country. However, the groups with the largest and smallest between-arm differences were assumed Group 1 and 4 respectively, according to the proportion of CRP+/MRI+ and male participants, but these were actually Group 1 and 3. Therefore, the presence of effect modifiers other than sex, CRP value, and MRI findings cannot be ruled out.

Table example 3-1 Percentage of participants with CRP+/MRI+ and males by group of participating countries (Study 301, ITT population)

Group of Participating Countries	m	CRP+/MRI+ (%)	Male (%)
All patients	375	26.7	42.1
Group 1 Canada, US, Denmark	95	33.7	51.6
Group 2 Sweden, Netherlands, France, Germany	73	30.1	42.5
Group 3 Brazil, Bulgaria, Belgium	79	26.2	46.6
Group 4 Japan, Switzerland, United Kingdom, Spain, China	103	24.1	38.0

m: number of patients

Figure example 3-4 ASAS response rate at Week 16 by group of participating countries (Study 301, ITT population)



n: number of responders, m: number of patients

5.3.2.2 Description example in M2.5.6.2

M2.5.6.2 Benefits

< Describe after the benefit claim based on overall study results >

Based on the results of PPK analysis of Japanese and foreign early clinical studies and multi-regional phase III study, Study 301, no intrinsic or extrinsic ethnic factors that have a clinically significant effect on PK were identified.

In Study 301, the ASAS response rate at Week 16 was analyzed by age, sex, race, body weight, CRP level, and the presence or absence of inflammatory signs on MRI findings. The ASAS40 response rate at Week 16 was higher in the test drug arm than in the control arm in all subgroups, although only limited number of subjects were aged ≥ 65 years or non-white, suggesting that the benefit of the test drug can be expected regardless of these patient background factors.

CRP levels or the presence or absence of inflammatory signs on MRI and sex were identified as effect modifiers of the test drug based on literature reports and analyses of the ASAS40 response rates at Week 16 in Study 301 by patient characteristics and by patient group. In Study 301, the ASAS40 response rate at Week 16 was evaluated in 6 groups of subjects stratified by the presence or absence of inflammatory signs based on CRP or MRI findings (CRP+/MRI+, CRP+/MRI-, CRP-/MRI+) and sex (male, female). As a result, the difference between groups

suggested that males compared to females and patients with both inflammatory signs of CRP level and MRI image findings compared to those with only one of them may benefit more from test drug. However, since the ASAS40 response rate at Week 16 in all groups was higher in the test drug arm than in the control arm, the benefit of the test drug can be expected regardless of the combination of these factors.

The number of subjects enrolled in Study 301 was small in many countries, making it difficult to evaluate the efficacy in each country. However, the results of the test of interaction in the ASAS40 response rate at Week 16 did not deny the consistency of efficacy among the countries that participated in the study. In addition, the ASAS40 response rate at Week 16 was higher in the test drug arm than in the control arm for all groups when participating countries were grouped into 4 groups by the proportion of male subjects and CRP+/MRI+ subjects. Although the presence or absence of inflammatory signs in CRP or MRI findings and the presence of effect modifiers other than sex cannot be ruled out, the between-arm differences between the groups were comparable, and therefore it is considered that the results in the overall population are applicable to each participating country. Therefore, the test drug is expected to provide consistent benefits across countries.

The number of Japanese subjects enrolled in Study 301 was as small as 8 (2 subjects in the test drug arm and 6 subjects in the control arm), and the efficacy of the test drug relative to control in the ASAS40 response rate at Week 16 was not demonstrated. However, it is not known that there are racial or regional differences in the distribution of patients on CRP level or the presence or absence of inflammatory signs in the MRI image findings and sex, which are identified as effect modifiers of the test drug. Thus, the results of the overall population were applicable to each participating country based on the above considerations. Therefore, the benefits of the test drug observed in the overall population can also be expected in Japanese patients.

5.3.3 Summary of the Five Perspectives

Biological Plausibility	It has been reported in the published literature that the CRP level or the presence or absence of inflammatory signs in the MRI image findings and sex that suggested the influence on the efficacy of the test drug are prognostic factors of axial spondyloarthritis and suggested the influence on drug response.
Internal Consistency	An analysis of subgroups in Study 301 based on the identified effect modifiers (CRP level or presence/absence of inflammatory signs on MRI image findings and sex) suggested that patients with both inflammatory signs of CRP level and MRI image findings compared to patients with only one of them and males compared to females might derive greater benefit from the test drug. On the other hand, analysis by country group in Study 301 based on the identified effect modifiers did not show the expected order of the between-arm difference.
External Consistency	It has been reported in the published literature that CRP levels or the presence or absence of inflammatory signs in MRI imaging findings and sex that suggested the influence on the efficacy of test drug suggest the influence on drug response.
Statistical Uncertainty	In some countries, the number of participants was small, and the estimates varied widely.
Clinical Significance	It is considered that there is no difference between Japan and overseas in the medical environment such as diagnosis and treatment of axial spondyloarthritis. There is no known racial or regional difference in the distribution of patients on CRP level or the presence/absence of inflammatory signs based on MRI image

	findings and sex, which are identified as effect modifiers of test drug. When countries were grouped into 4 groups based on identified effect modifiers in Study 301, the between-arm difference in the ASAS40 response rate at Week 16 was similar among the groups.

6 Summary

This document describes points to consider when preparing the CTD based on the multifaceted and structured evaluation of MRCT results in accordance with the principles of ICH E17 and the 3-layer approach (Komiyama et al. 2013) proposed by Komiyama et al.

There has been an increase in the use of MRCTs including countries other than developed countries and global development including MRCTs, with an aim to provide patients worldwide with early access to necessary drugs. ICH E17 "General Principles for Planning and Design of Multi-regional Clinical Trials" was implemented in 2018, and conducting MRCTs in accordance with ICH E17 has an advantage that the benefit-risk of drugs can be evaluated from various perspectives based on various racial and ethnic factors. However, at present, most of CTDs are based on the concept of "local first," which focuses on results in the country/region of application and then evaluates similarity to other regions or the overall study. This assessment approach has worked to some extent if the number of participants in the country/region of application is adequate to assess similarity. However, since the overall number of participants in the MRCT will be determined depend on the objective of the study, it is expected that the number of participants in each country/region participating in the MRCT will be inevitably small. In addition, if there are many countries/regions participating in the MRCT, the number of participants in each country/region may be even smaller, which increases statistical uncertainty due to small numbers. The attempt to perform subgroup analyses with effect modifiers further increases the statistical uncertainty.

What was emphasized in this document is not to debate "local results" versus "overall MRCT results," but first review overall MRCT data from the perspective of exploring effect modifiers or potential candidates for them, and to use MRCT results to help predict and understand therapeutic effects of drugs. In addition, based on this consideration, there was an emphasis on the importance of characterizing the patient population in the country/region of application with effect modifiers and estimate the treatment effect using them. The 3-layer approach was described, which shows the flow of evaluating the MRCT results; it was encouraged to put this approach into practice. The incorporation of the 3-layer approach in the CTD is expected to result in positive changes in regulatory submissions documents and discussions between industry and regulatory agencies during the review process. Enrichment of evaluation up to Layer 2 using the overall MRCT data means to organize knowledge of what factors (including effect modifiers, or country/region as the phenotype of the potential effect modifier) may affect the treatment effect. The examination results up to Layer 2 should be useful in all countries in the world, and for this purpose, it will be necessary to enhance Layer 2, which is not intended to apply to specific countries/regions, in the CTD prepared by the global team. Each country/region of application can then consider their country/region's Layer 3 based on the globally shared Layer 2 findings. This may reduce the amount of work that needs to done individually in each country/region of application, increase the common part of the CTD in each country/region of application, better organize discussions with regulatory authorities during the review phase, and create a possibility of providing each authority with basically the same story. It may also be possible for one country/region of application to refer to the evaluation in other countries/regions and brush up the discussion. Although the conclusion based on "local results versus overall MRCT results" might not be immediately applicable to other countries/regions, discussion around effect modifiers could facilitate application of MRCT results to local patient populations, with effect modifiers becoming a "common language around the world." When

the benefit/risk of drugs is evaluated from a common perspective worldwide, evidence will be accumulated not only at the development stage but also at the post-marketing drug fostering stage while cooperating worldwide. In the process, it may be possible to discuss Japan's contributions.

In this document, we introduced many examples of analysis using the 3-layer approach from the five perspectives of ICH E17. As presented, the analytical methods themselves are not new, and many conventional methods are included. It is expected that the planning and interpretation of MRCT results will be different from before by setting the framework of 3-layer approach, and the five perspectives in ICH E17 as the core, and planning and implementing an analysis method that enables to discuss the benefit-risk in the country/region of application after making maximum use of available information.

We hope that this document will contribute to the interpretation of MRCT results and the dissemination of CTD preparation based on the principles in ICH E17.

7 References

Baselga J, Cortés J, Kim SB, et al. (2012) Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. N Engl J Med; 366(2):109-19.

Bonaca MP, Bhatt DL, Cohen M, et al. (2015) Long-term use of ticagrelor in patients with prior myocardial infarction. N Engl J Med; 372(19):1791-800.

Brown DM, Emanuelli A, Bandello F, et al. (2022) KESTREL and KITE: 52-Week Results From Two Phase III Pivotal Trials of Brolucizumab for Diabetic Macular Edema. Am J Ophthalmol; 238:157-72.

KJ Carroll and TR Fleming (2013) Statistical Evaluation and Analysis of Regional Interactions: The PLATO Trial Case Study, Stat Biopharm Res, 5:2, 91 91-101.

Klepper MJ and Cobert B (2012) Drug Safety Data: How to Analyze, Summarize and Interpret to Determine Risk, Scientist Press Co., Ltd.

Komiyama O, Hiro S, Isogawa N, et al. (2013) Evaluation of Data for Multi-Regional Trials: A Three-Layer Approach. Applied clinical trials, vol 22, Issue 11.

Swain SM, Kim SB, Cortés J, et al. (2013) Pertuzumab, trastuzumab, and docetaxel for HER2-positive metastatic breast cancer (CLEOPATRA study): overall survival results from a randomised, double-blind, placebo-controlled, phase 3 study. Lancet Oncol; 14(6):461-71.

Uyama Y, Shibata T, Nagai N, et al. (2005) Successful bridging strategy based on ICH E5 guideline for drugs approved in Japan. CLINICAL PHARMACOLOGY & THERAPEUTICS; 78(2):102-13.

Warner TD, Nylander S, Whatling C. (2011) Anti-platelet therapy: cyclo-oxygenase inhibition and the use of aspirin with particular regard to dual anti-platelet therapy. Br J Clin Pharmacol; 72(4):619-33.

Yoshida K, Awaji N, Takenouchi K, et al. (2015) How Should a Globalized CTD Be Created? An Introduction to the Japanese 3-Layer Approach. Ther Innov Regul Sci; 49(1):175-80.

ICH E5 Ethnic Factors in the Acceptability of Foreign Clinical Data (1998) (https://www.pmda.go.jp/files/000156571.pdf, Last Accessed July 21, 2023)

ICH E5 Q&A on "Ethnic Factors in the Acceptability of Foreign Clinical Data" (Part 2)(2006) (https://www.pmda.go.jp/files/000156024.pdf, Last Accessed July 21, 2023)

ICH E17 General Principles for Planning and Design of Multinational Clinical Trials (2018) (https://www.pmda.go.jp/files/000224557.pdf, Last Accessed July 21, 2023)

ICH E17 General Principles for Planning and Design of Global Clinical Trials "training materials, etc. (https://www.pmda.go.jp/int-activities/int-harmony/ich/0022.html, Final Access Confirmation July 21, 2023)

Review Report of Enfortumab vedotin (Genetical Recombination):Unresectable Urothelial Cancer That Progressed After Cancer Chemotherapy (2021) (https://www.pmda.go.jp/drugs/2021/P20211011002/800126000_30300AMX00454000_A100_1.pdf, Last Access Confirmed July 21, 2023)

General principles on Multi-Regional Clinical Trials (PFSB/ELD Notification No. 0928010) (2007) (http://wwwhourei.mhlw.go.jp/cgi-bin/t_document.cgi?MODE=tsuchi& (japal.org), last access confirmed July 21, 2023)

The Practice of 3-layer approach that is the Methods for Examining Ethnic Factors in Multi-Regional Clinical Trials (Japan Pharmaceutical Manufacturers Association 2018) (https://www.jpma.or.jp/information/evaluation/results/allotment/lofurc000000a0k7-att/3_layer_approach_final.pdf, Last Access Confirmed July 21, 2023)

Drug Evaluation From Now On -What is the value of local data in the global clinical data package? - (Japan Pharmaceutical Manufacturers Association 2022) (https://www.jpma.or.jp/information/evaluation/results/allotment/gbkspa0000001336-att/DS_202207_e17etc.pdf, Last Access checked July 21, 2023)

2nd ICH E17 Workshop (2022) (https://www.pmda.go.jp/files/000247710.pdf,https://www.pmda.go.jp/files/000247711.pdf,https://www.pmda.go.jp/files/000247712.pdf,https://www.pmda.go.jp/files/000247713.pdf,https://www.pmda.go.jp/files/000247714.pdf, Final Access Confirmation July 21, 2023)

Summary of Application data on Dabigatran Etexilate Methanesulfonate (2011) Prevention of ischemic stroke and systemic embolism in patients with non-valvular atrial fibrillation (https://www.pmda.go.jp/drugs/2011/P201100019/index.html, Last access checked July 21, 2023)

Ticagrelor US Review Related Information (2010) (https://wayback.archive-it.org/7993/20170405212359/https:/www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/CardiovascularandRenalDrugsAdvisoryCommittee/UCM221383.pdf,https://wayback.archive-it.org/7993/20170405212347/https:/www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/CardiovascularandRenalDrugsAdvisoryCommittee/UCM220197.pdf, Final Access Confirmation July 21, 2023)

Review Report of Finerenone: Chronic kidney disease complicated by type 2 diabetes mellitus (2022) (https://www.pmda.go.jp/drugs/2022/P20220317001/630004000_30400AMX00176_A100_1.pdf, Last access checked July 21, 2023)

FY 2022 Data Science Subcommittee Task Force 2-2

Osamu Komiyama (Pfizer R&D G.K.): Deputy Review Director

Yuko Asahara (Novartis Pharma K.K.): Leader

Masako Aoya (AbbVie G.K.)

Yasuko Kasai (Novartis pharma K.K.)

Yuki Saito (Janssen Pharmaceutical K.K.)

Tomoko Sakuma (Bristol Myers Squibb K.K.)

Yuji Sasagawa (Meiji Seika Pharma Co., Ltd.)

Noriko Seko (Novartis Pharma K.K.)

Masaya Marumo (Kissei Pharmaceutical Co., Ltd.)