Challenges Faced Conducting Interim Biomarker Exploration in a Phase II Study

○Hiroki Yoshida[1], Hideharu Yamamoto[1], Yasuo Sugitani[1,2] and Miwa Nakano[1]

[1]: Chugai Pharmaceutical CO., Ltd.

[2]: Genentech Inc.

Background：Purposes of Biomarker（BM） development and requirements from Agencies

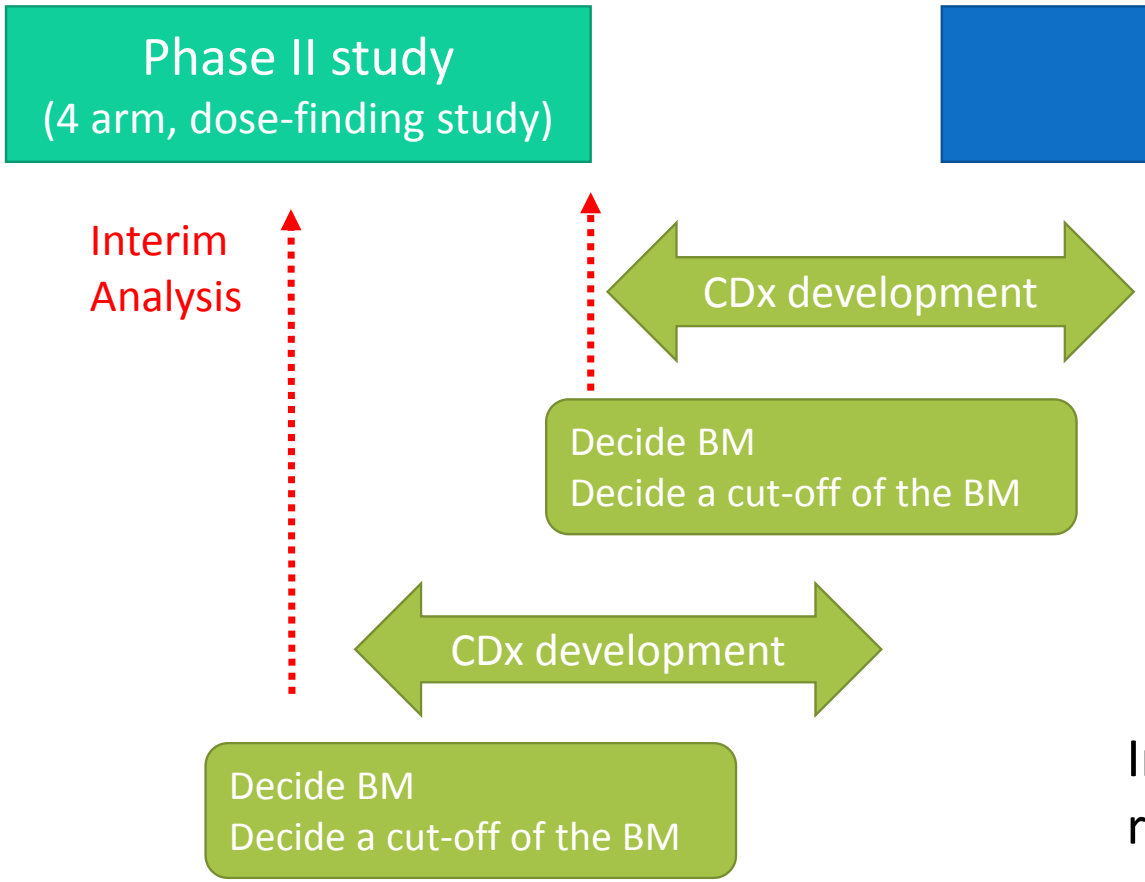
Purposes of BM development

- To select population which has much efficacy to differentiate the drug from competitors
- To select population which has much efficacy decrease the number of patients in P3 study and budget of P3.

Requirements to develop BM from Agencies (EMA guideline)

- BM should be assessed by clinically and statistically valid method.
- BM should be assessed in a prospective manner in pivotal studies.

Development status of a drug in Chugai



In this case, we can't evaluate the relationship between efficacy and BM in a phase III study in a prospective manner because CDx is not prepared by phase III start.

In this case, we can evaluate the relationship in a prospective manner!

Purposes of this presentation

Implementation structure of the interim analysis

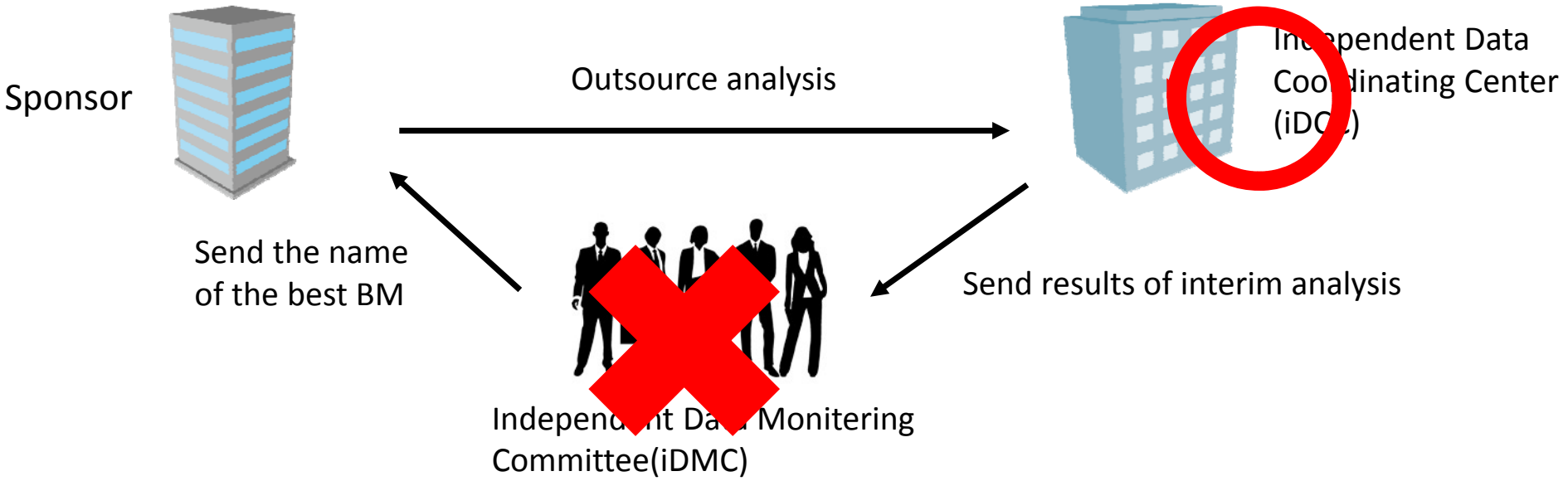
- To share challenges to keep acceptable integrity of a blinded study for conducting an interim analysis for Biomarker (BM) and how to overcome the issues.

Analysis to select BM

- To share what kind of analysis was used for BM selection

Issue of Implementation structure :
Who should analyze, evaluate BM and make decision in the interim analysis?

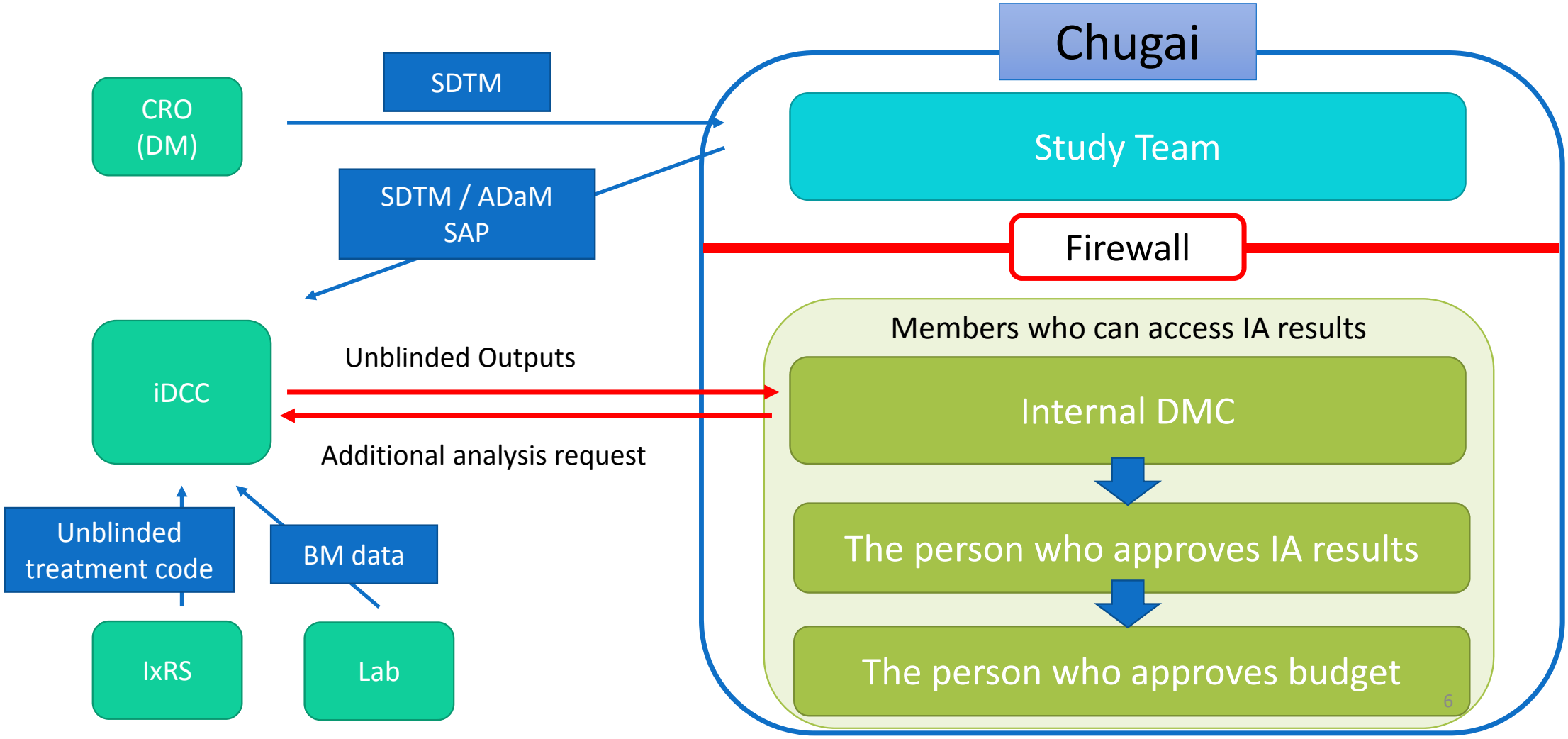
- To outsource all tasks of interim analysis is ideal.
  - We can keep acceptable integrity of a blinded study with that.



We need to consider several things totally to make decision for developing CDx.

We established an internal DMC since BM exploration required a flexible approach.

Biomarker Interim Analysis Organization



Minimalize the members who know results of interim analysis

- Only seven members in internal DMC to evaluate results.

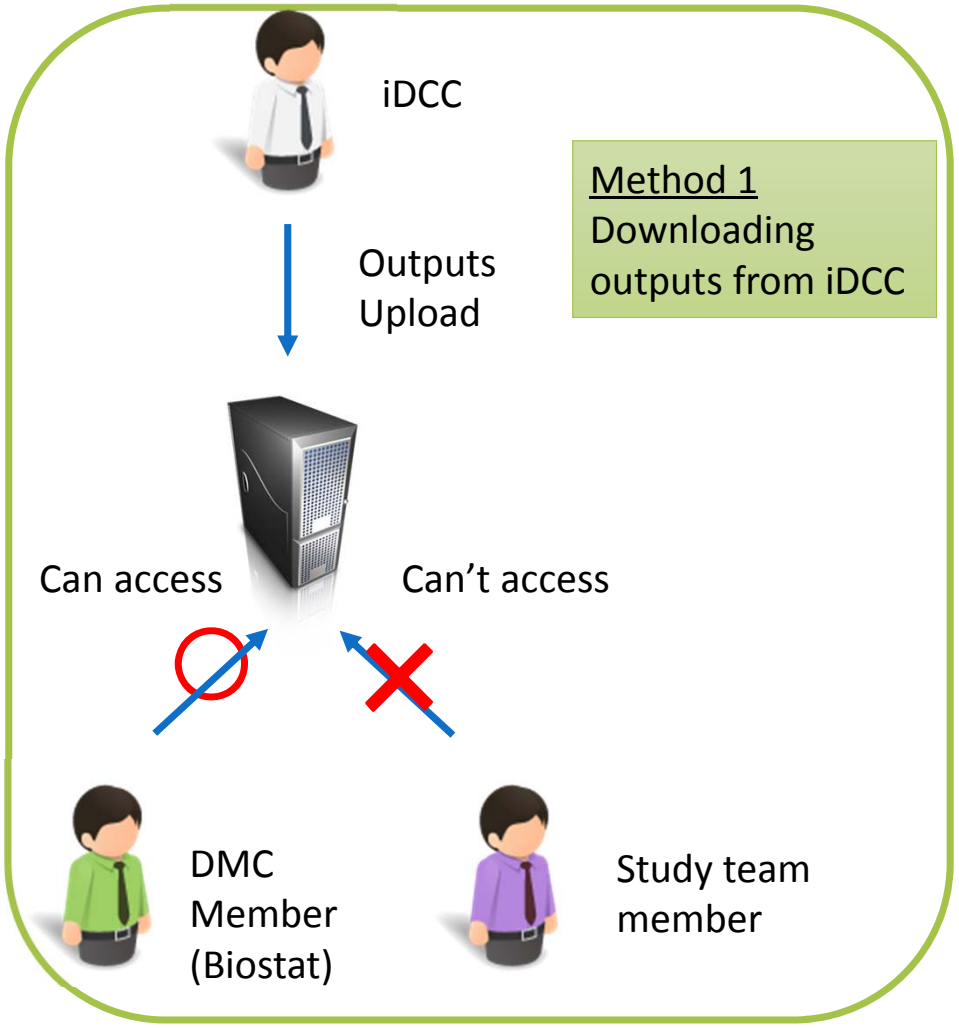
Clinical Pharmacology (2 members) | Clinical Science | Personalized Healthcare | Discovery Research | Biostat (2 members)

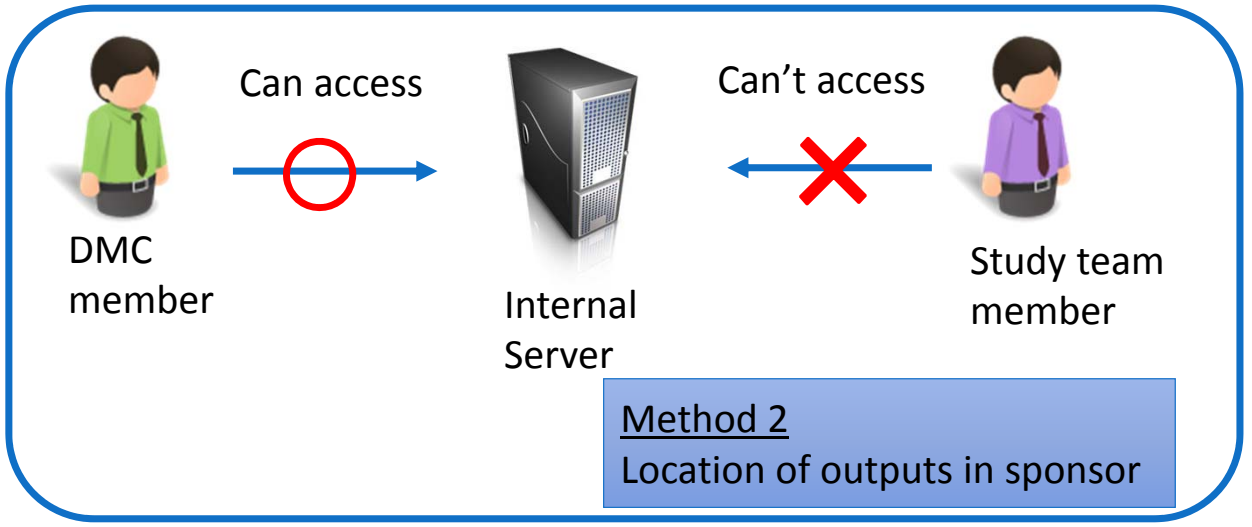
- Only two members to make decision for BM development
  - No board held and a chairperson can approve

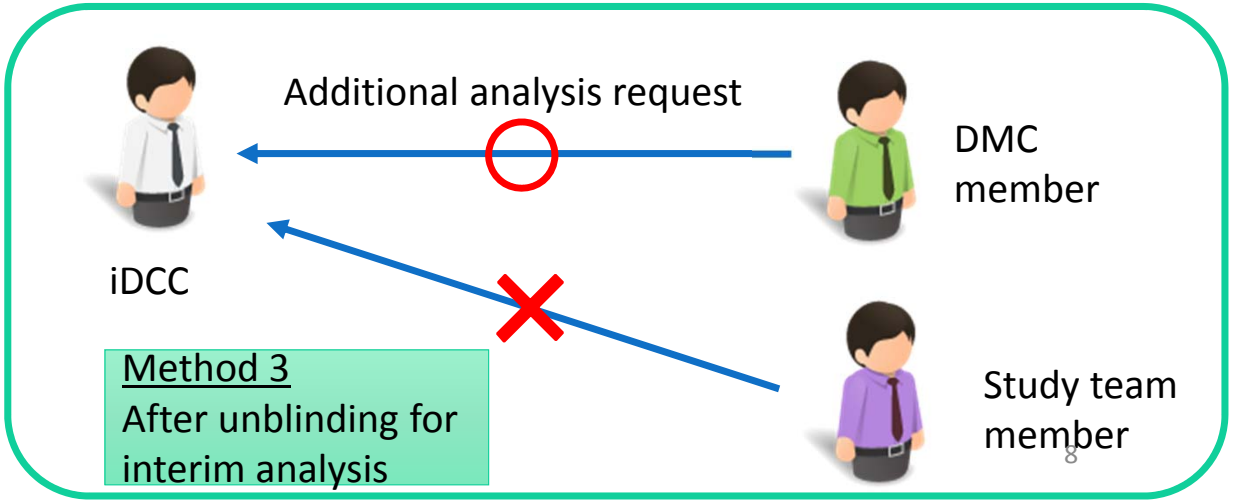
Approve IA results

Approve budget

Methods for firewall







Summary 1

- We carried out an interim analysis for BM in a P2 study to develop CDx which can be used for P3 study P3.
- We established an internal DMC since BM exploration required a flexible approach.
- The member of the internal DMC were independent from study team.
- Several methods for constructing firewall in sponsor was taken.

Purposes of this presentation

Implementation structure of the interim analysis

- To share challenges to keep acceptable integrity of a blinded study for conducting an interim analysis for Biomarker (BM) and how to overcome the issues.

Analysis to select BM

- To share what kind of analysis was used for BM selection

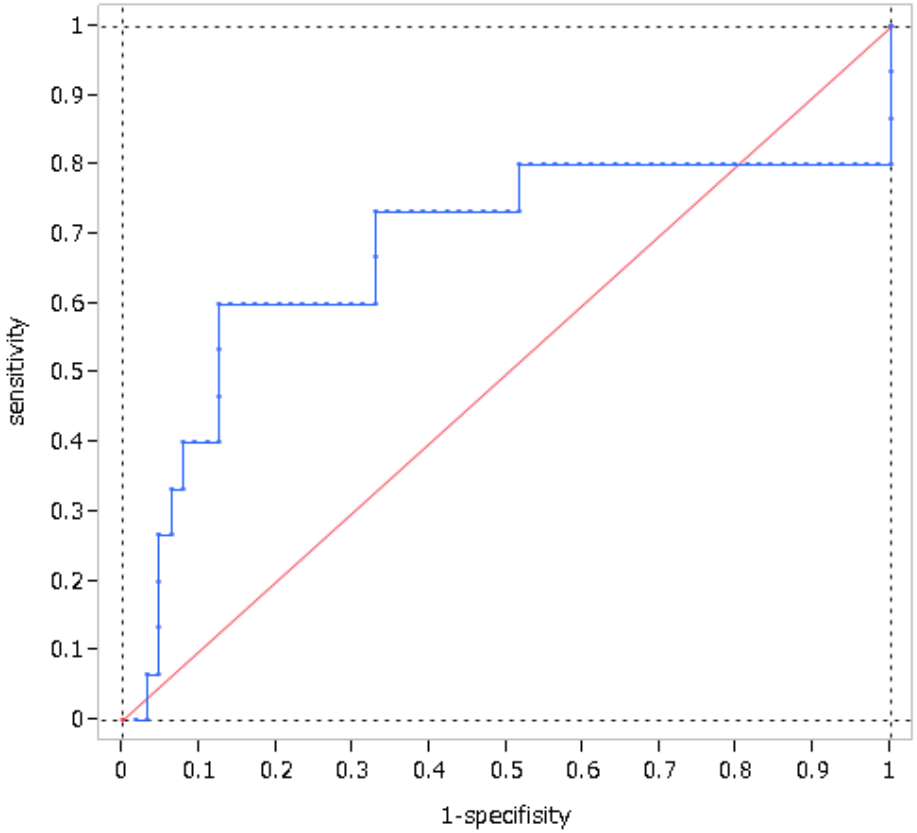
What kind of analysis was adopted for BM selection?

Analysis	Why did we adopt the analysis?
Summary of Demographics and Baseline Characteristics	To confirm difference between groups
Summary of efficacy parameters	To confirm the product has good efficacy for patients or not. (If not, it makes no sense to develop CDx.) To confirm there are both responders and non-responders. (If not , it makes no sense to develop CDx.)
Relationship between BM and efficacy	To confirm there are some BM relating to efficacy or not.
AUC of ROC curve of test set	To confirm we should develop a CDx for the BM.

What is ROC curve and AUC?

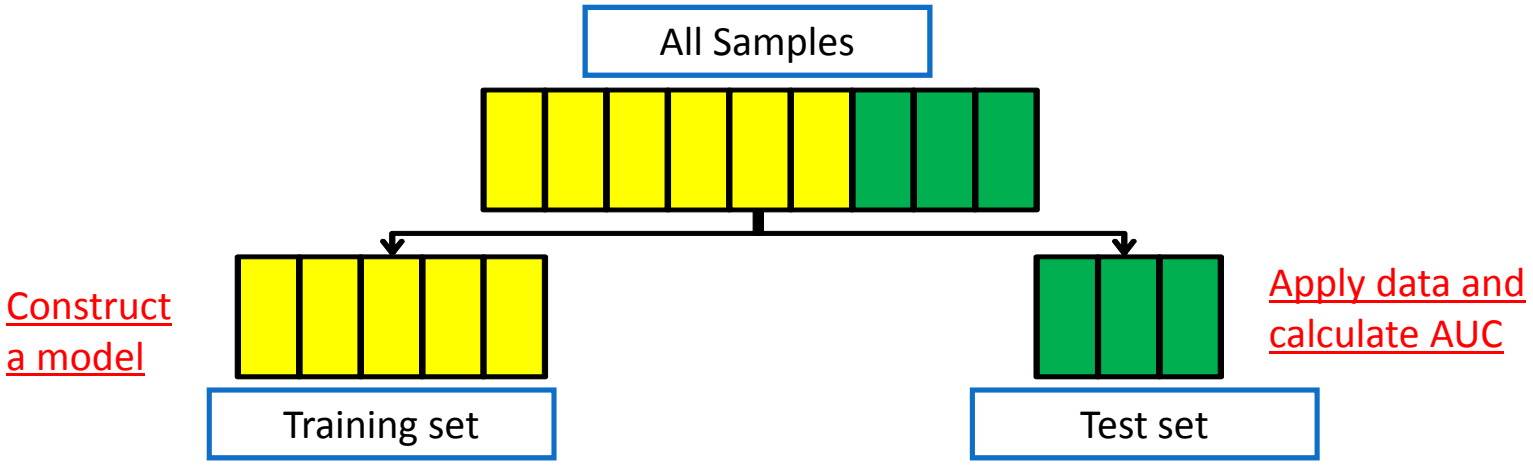
- ROC curve is the curve we can understand relationship between sensitivity and specificity visually .

- The Area Under the Curve (of ROC).
  - If BM doesn't relate to efficacy, AUC is around 0.5.
  - We considered that BM with AUC >= 0.8 should be selected.



Method to evaluate the ability to determine responder and non-responder without over-fitting.

- We validated the capacity to divide all samples into Training set and Test set to avoid over- fitting.
  - Validation in the next study is ideal, but not realistic.



- We adopted logistic model because we divided efficacy parameters into Responder and Non-responder.

Two issues for BM assessment

Sufficient number of patients for BM evaluation?

- There were **only 50 patients per group** for evaluating BM at the interim analysis.

Issue of multiplicity by evaluating many BM candidates

- There were more than 40 BM candidates before the interim analysis so selected BM might be false-positive.

We conducted simulations to evaluate above issues.

Simulation 1:
For confirming error and power with 50 patients.

- Setting

  * The number of simulation : 10,000

  - N = 50Responder (R) : Non-responder (NR) = 1 : 1
  - Training : Test = 3: 2

	Responder	Non-responder	Total
Training	15	15	30
Test	10	10	20
Total	25	25	50

  - One BM.
  - Distribution of BM was assumed normal distribution referring to the results of phase I study
    - When considering error, R and NR follow same distribution.(Null hypothesis)
    - When considering power, R and NR follow each distribution. (Alternative hypothesis)

Result of simulation 1

- Error : Proportion of the results beyond AUC=0.7 under null hypothesis in 10,000 simulations.
- Power : Proportion of the results beyond AUC=0.7 under alternative hypothesis in 10,000 simulations.

Number of patients (Training:Test)	Error	Power
50 (30:20)	6.4%	97.8%

Simulation 2:
For confirming increase of error due to multiplicity

- Setting

* The number of simulation : 10,000

  - N = 50
  - Responder (R) : Non-responder (NR) = 1 : 1
  - Training : Test = 3: 2

	Responder	Non-responder	Total
Training	15	15	30
Test	10	10	20
Total	25	25	50

  - 40 BMs
  - When considering error, R and NR follow same distribution. (Null hypothesis)

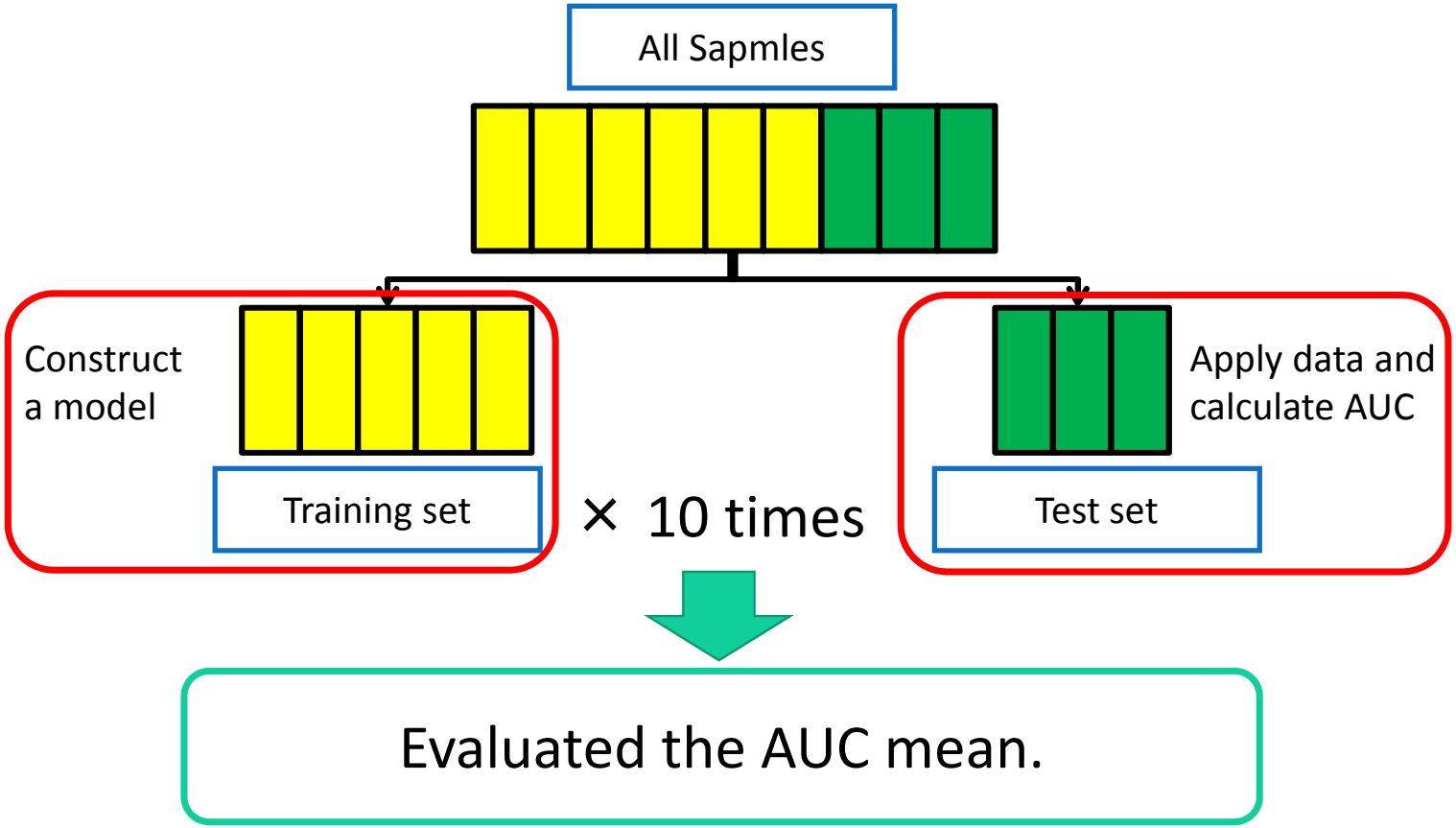
Results of simulation 2

Threshold	Proportion beyond threshold at least one BM of 40 BMs (error) (Under null hypothesis)
AUC=0.7	92%
AUC=0.75	67%
AUC=0.8	33%

When the number of BM is only one, we can keep error under 7%.
However, when the number of BM is 40, proportion beyond threshold (AUC=0.7) at least one BM increase to 92% (AUC=0.7).

How did we avoid 92% error?

- We evaluated AUC mean of test set when we re-sampled test set 10 times randomly.



Result of error for the AUC mean

	Proportion beyond threshold at least one BM of 40 BMs (error) (under null hypothesis)	
Threshold	**10 times sampling**	**100 times sampling**
AUC=0.7	54%	38%
AUC=0.75	12%	8%
AUC=0.8	2%	1%

We can decrease error by this method.
(10 times re-sampling and threshold with AUC=0.8)

Summary of simulation

- We overcame the issue of multiplicity with the method re-sampling test set 10 times randomly.

	The % of one or more of 40 BM beyond threshold (under null hypothesis)		
Threshold	**One time sampling**	**10 times sampling**	**100 times sampling**
0.7	92%	54%	38%
0.75	67%	12%	8%
0.8	33%	2%	1%

Summary 2

- We selected AUC of test set for evaluating BM.
- We conducted simulations to overcome issues of the number of patients and multiplicity of BMs.
- Error caused by multiplicity was decreased by the method of 10 times re-sampling.

Summary of this presentation

- We shared challenges to keep acceptable integrity of a blinded study for conducting an interim analysis for Biomarker (BM) and how to overcome the issues.

- We shared what kind of analysis was adapted for BM selection

- We shared error caused by multiplicity was decreased by the method of 10 times.