

# 第6回 DSRT

## ベイズ統計学を利用して、既存情報を臨床試験 にどう有効活用するか？

本資料に含まれる議論結果については、議論の中で挙げた意見を網羅的にまとめたものであり、コンセンサスの得られた見解ではないことにご留意ください。

# 本テーマの目的

- 臨床開発の効率化のため、co-data（ヒストリカルデータ、同時データ）の有効活用が注目されている。昨今、ベイズ統計を利用したco-dataの利用法が提案されている（例えばNeuenschwander et al. 2016）。ベイズ統計学を利用することで、試験間あるいはregion（地域）間の違いを考慮しながら（dynamicに）互いの情報を利用し合うことができる。この特徴を活かすことで、より効率的な臨床開発が可能になることが期待される。
- そこで、今回は、今後の臨床開発の効率化につながることを期待されるベイズ流co-data活用場面の例として、国際共同治験（Multi-Regional Clinical Trials; MRCT）におけるregion別の有効性評価及び第I相試験から第III相試験を通じた包括的な有効性評価を取り上げ、適切な利用方法、及び利用する際に注意すべき点について、参加者で議論する。

# Co-dataとは何か？

有効活用することで、  
より効率的な新薬  
開発が可能に！！

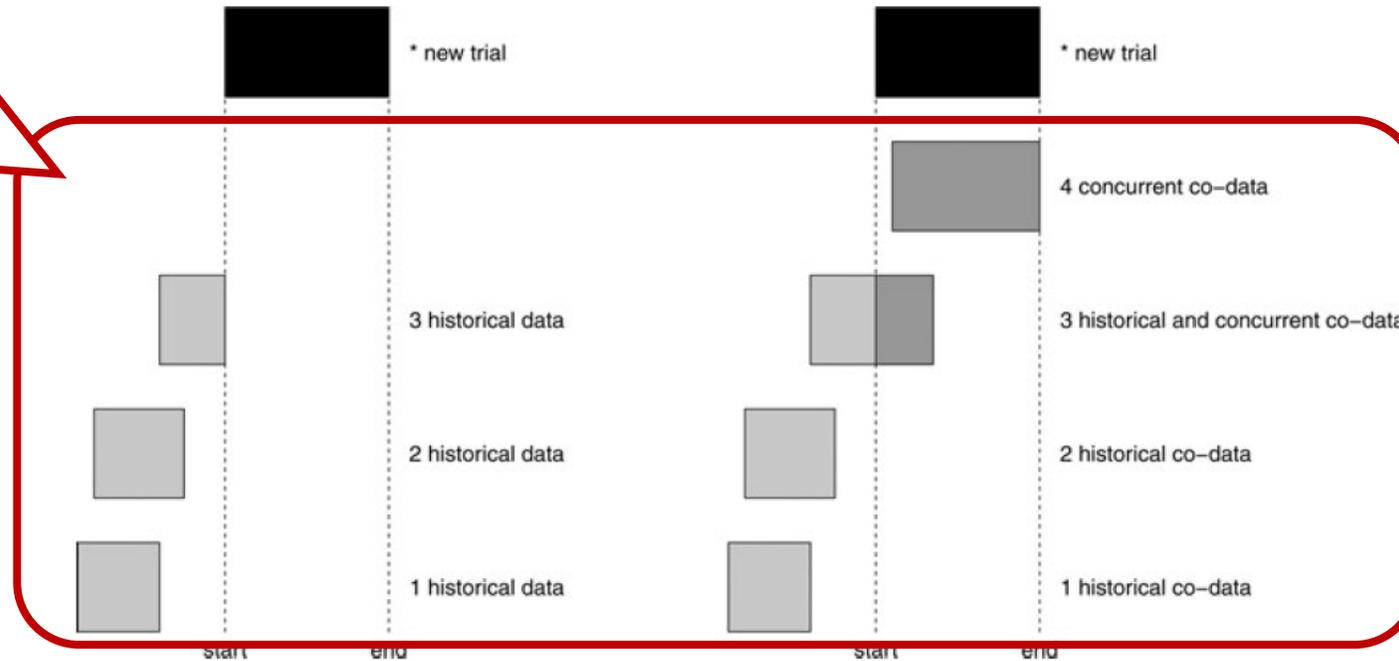


Figure 1. Two cases: new trial with historical co-data (left panel), and with historical and concurrent co-data (right panel).

# 臨床試験におけるヒストリカルデータの利用に関する 主な規制当局からの言及 (Lim et al. 2018)

Table 1. Key Regulatory References.

Reference Document	Key Message(s)
ICH E4: Dose Response Information to Support Drug Registration <sup>9</sup>	“Agencies should also be open to the use of various statistical and pharmacometric techniques such as Bayesian and population methods, modeling, and pharmacokinetic-pharmacodynamic approaches.”
ICH E9: Statistical Principles for Clinical Trials <sup>10</sup>	“The use of Bayesian and other approaches may be considered when the reasons for their use are clear and when the resulting conclusions are sufficiently robust.”
ICH E10: Choice of Control Group and Related Issues in Clinical Trials <sup>11</sup>	Guideline expresses a major concern about only using historical controls (ie, the inability to control bias), but also describes the usefulness of such controls under certain scenarios. Guideline describes situations where appropriately and carefully chosen historical controls are more persuasive and potentially less biased.
FDA: The Use of Bayesian Statistics in Medical Device Clinical Trials <sup>12</sup>	Guidance states that in some circumstances “the prior information for a device may be a justification for a smaller-sized or shorter-duration pivotal trial.”
EMA: Concept Paper on Extrapolation of Efficacy and Safety in Medicine Development <sup>13</sup>	Paper proposes a framework to establish a systematic approach for extrapolation of efficacy and safety data from a source population to a target population.
EMA: Guideline on Clinical Trials in Small Populations <sup>14</sup>	A Bayesian methodology with an informative prior built on historical data may be suitable. Recommends sensitivity analyses for the choice of prior. “Under exceptional circumstances,” historical controls with no concurrent control may be acceptable.
European Commission: Ethical Considerations for Clinical Trials on Medicinal Products Conducted with the Paediatric Population <sup>15</sup>	“Adaptive, Bayesian, or other designs may be used to minimize the size of the trial.”

# ヒストリカルデータを用いて事前分布を設定する際のオプション (Lim et al. 2018)

Table 2. Options for Specifying a Prior Based on Historical Data.

Assumed Relationship Between $p_H$ & $p$	Details
1. Equal	Assumes differences between observed historical & concurrent control response rates are solely attributable to sampling variation. Equivalent to pooling historical & concurrent controls.
2. Functional Dependence	Assumes differences between historical and concurrent control response rates can be explained by known covariates. Requires estimate of covariate-response relationship (eg, from historical data) which can be used to construct a predicted prior distribution for $p$ based on historical control response rate and observed covariates for historical and concurrent controls.
3. Equal but Discounted	Assumes $p_H = p$ as in option 1, but discounts the historical information by inflating the variance of the historical prior (eg, using power prior). <sup>21</sup> Can also be thought of as reducing the effective sample size of the historical controls on which the prior is based. Amount of discounting is subjective and has no operational interpretation.
4. Biased	Assumes concurrent control response is a biased (shifted or rescaled) version of the historical response (ie, $p = p_H + \delta$ or $p = p_H \times \delta$ ). Similar to functional dependence (option 2), except that the precise form of dependence is unknown. Therefore, the dependence is captured by a generic bias term $\delta$ . Prior for $p$ is constructed by combining the historical prior for $p_H$ with a prior for $\delta$ which is typically chosen to reflect judgments about the relevance and quality of the historical study to the current setting. <sup>22</sup>
5. Exchangeable	Assumes $p_H$ and $p$ are “similar” (ie, assumes a distribution across studies with parameter $\sigma^2$ that reflects heterogeneity between historical and concurrent control response rates). Equivalent to a random effects meta-analysis of the historical and current trials.

# ベイズ流枠組みにおける動的借用アプローチ (Lim et al. 2018, 武田他 2015)

手法	詳細
Commensurate priors	ヒストリカルデータと新規試験データの共通性 (commensurability) に基づいてヒストリカルデータから得られる情報量を適応的に調整する方法
Power priors	ヒストリカルデータの尤度をべき乗パラメータ (power parameter) $a^0 \in [0, 1]$ でべき乗することにより、ヒストリカルデータの情報量を割り引く方法 既知の定数を使う方法 (conditional power prior) とべき乗パラメータに事前分布を設定する方法 (joint power prior, modified power prior) がある。
Meta-analytic-predictive priors (MAP priors)	新規試験の対照治療群のパラメータと既存試験の対照治療群のパラメータが同じ分布からのランダムサンプルとみなす階層モデルを用いる方法 情報を借用する程度は試験間の不均質性 (heterogeneity) に依存
Robust MAP priors	新規試験の対照治療群のパラメータと既存試験の対照治療群のパラメータに交換可能性 (exchangeability) を仮定するが、交換可能性が成り立たない可能性も考慮した方法。事前分布はMAP priorとvague priorの混合分布

議題1:  
MRCTにおけるregion別の有効性評価

# 適用場面1:国際共同治験（MRCT）における 国・地域毎の有効性評価

- ICH-E17において、「個々の地域における治療効果とそれらの不確実性の程度を推定し、報告するための適切な統計手法を、治験実施計画書の統計解析の項に記載すべき」とされている。
- 治療効果の報告に関心のある各国・地域において、症例数が非常に少なく、単純な部分集団解析に基づき有効性を判断することは困難な場合がある。
- そのような場合における解決案の一つとして、Bayesian shrinkage estimator を用いることが提案されている（例えば、Quan et al. (2013) 等）。

# Quan et al. (2013) の提案の概要

- 国・地域における治療効果 $\delta_i$  を共通の母集団から抽出された標本であると考ええる。
  - $\delta_i$  は、全国・地域の平均治療効果 $\delta$  と地域間変動 $\tau^2$  を母数とする正規分布に従うものとする ( $\delta_i \sim N(\delta, \tau^2)$ )。
- 上記の仮定の下で、Bayes 流統計手法を利用し、得られた臨床試験データに基づき、 $\delta_i$  を推定したものが、Bayes流縮小推定値となる。

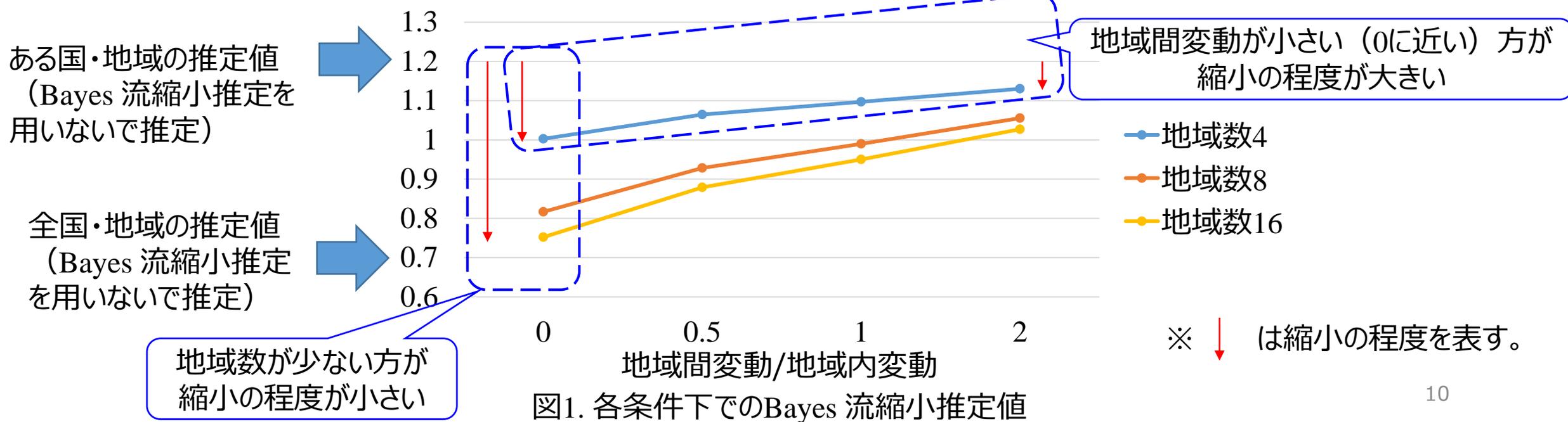
$$\hat{\delta}'_i = \hat{\delta}_i + \frac{s-3}{s-1} \frac{\hat{\omega}_i^2}{\hat{\tau}^2 + \hat{\omega}_i^2} (\hat{\delta} - \hat{\delta}_i)$$

全体集団の結果への縮小の程度

$\hat{\delta}_i$  : Bayes 流統計手法を用いずに推定した国・地域の治療効果の推定値、  
 $s$ : 国・地域数、 $\hat{\omega}_i^2$ : 国・地域内におけるデータのばらつきの推定値 (地域内変動)、  
 $\hat{\tau}^2$ : 地域間変動の推定値、 $\hat{\delta}$ : 全国・地域の治療効果の推定値 (変量効果モデルを利用)

# Quan et al. (2013) の方法の特徴

- 国・地域における治療効果 $\delta_i$  を共通の母集団から抽出された標本であると考える。
- 上記の仮定の下で、Bayes 流統計手法を利用し、得られた臨床試験データに基づき、 $\delta_i$  を推定したものが、Bayes 流縮小推定値である。



# 適用例

- 3991例

**Table IV.** Data of mortality from MERIT-HF.

Country	Sample size		Number of events	
	Meto CR/XL	Placebo	Meto CR/XL	Placebo
Belgium	68	66	3	13
Czech Republic	123	124	9	17
Denmark/Finland	141/20	150/14	11/0	11/2
Germany	252	247	19	31
Hungary	211	212	16	29
Iceland	19	22	2	2
Norway	97	105	6	11
Poland	102	102	8	8
Sweden	39	46	2	9
The Netherland/Switzerland	278/21	270/21	14/0	25/1
UK	87	83	4	9
USA	532	539	51	49

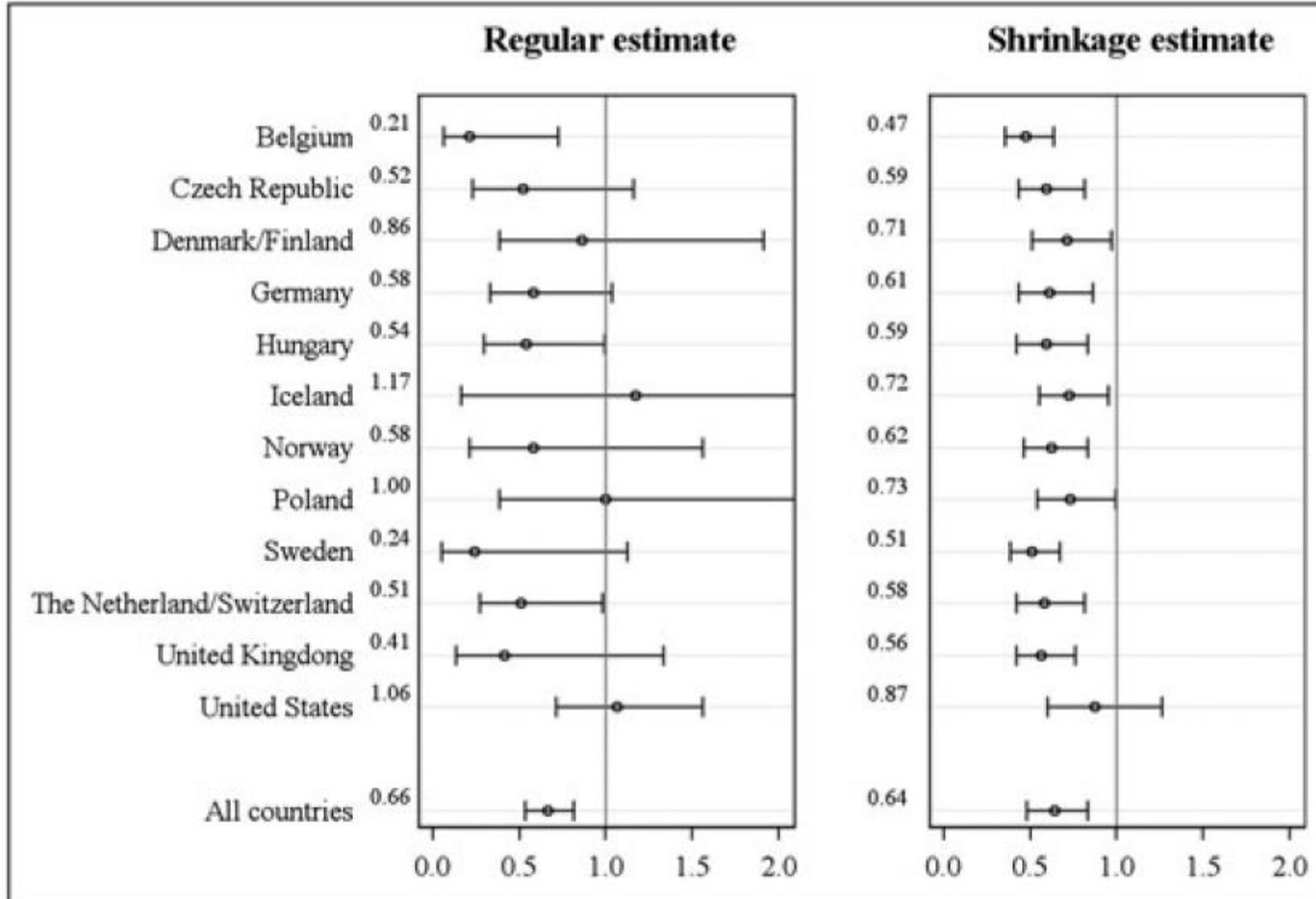
# 結果

- 全体集団で観察された治療効果：0.66 [0.53, 0.81]
- 国・地域と治療効果の交互作用：p値 = 0.22、 $\hat{\tau}^2 = 0.04$

**Table V.** Results on mortality for MERIT-HF.

Country	$e^{-\hat{\delta}_i}$ (95% CI)	$\hat{\delta}_i$	$\hat{\sigma}_i^2$	$e^{-\hat{\delta}'_i}$ (95% CI)	$e^{-\hat{\delta}'_i}$ (95% CI)
Belgium	0.21 (0.06, 0.72)	1.58	0.41	0.57 (0.43, 0.76)	0.47 (0.35, 0.63)
Czech Republic	0.52 (0.23, 1.16)	0.66	0.17	0.61 (0.44, 0.83)	0.59 (0.43, 0.81)
Denmark/Finland	0.86 (0.38, 1.91)	0.15	0.17	0.68 (0.49, 0.93)	0.71 (0.51, 0.97)
Germany	0.58 (0.33, 1.03)	0.54	0.08	0.62 (0.44, 0.87)	0.61 (0.43, 0.86)
Hungary	0.54 (0.29, 0.99)	0.62	0.10	0.60 (0.43, 0.84)	0.59 (0.42, 0.83)
Iceland	1.17 (0.16, 8.28)	-0.15	1.00	0.65 (0.50, 0.86)	0.72 (0.55, 0.95)
Norway	0.58 (0.21, 1.56)	0.55	0.26	0.63 (0.46, 0.85)	0.62 (0.46, 0.83)
Poland	1.00 (0.38, 2.66)	0.00	0.25	0.68 (0.50, 0.92)	0.73 (0.54, 0.99)
Sweden	0.24 (0.05, 1.12)	1.42	0.61	0.60 (0.45, 0.79)	0.51 (0.38, 0.67)
The Netherland/Switzerland	0.51 (0.27, 0.98)	0.67	0.11	0.60 (0.43, 0.83)	0.58 (0.42, 0.81)
UK	0.41 (0.13, 1.33)	0.89	0.36	0.61 (0.45, 0.81)	0.56 (0.42, 0.76)
USA	1.06 (0.71, 1.56)	-0.06	0.04	0.83 (0.57, 1.20)	0.87 (0.60, 1.26)

# 結果



Quan et al. (2013)

# 論点

- MRCTにおいて、各国・地域における治療効果の推定にQuan et al. (2013) の方法を利用することを考える場合、どういう検討をする必要があるだろうか？

例えば、

- 情報借用することの妥当性は？
  - 全地域に共通の分布を仮定することはリーズナブルなのか（Guo et al. (2016)の議論等）？
  - どんな情報を使ってどうやって説明するか？
- どんな状況でもQuan et al. (2013) の方法の利用を考えるべきか？
  - 提案法の懸念点はないか（性能の観点等）？

## 議論結果①

- 縮小推定を適用するにあたって考慮すべきと考える点
  - シミュレーション等による性能の確認
  - 縮小推定を用いる必要性（以下のような点を考える）
    - 国・地域ごとの部分集団解析結果の精度等
    - 国・地域の併合可能性
  - 縮小推定の位置づけ（主要な解析か補足的な解析か）
  - どこまでの内容をプロトコルに含めるべきか
  - 各国・地域の症例数設定

## 議論結果②

- 他国・地域の情報を借用することの妥当性について
  - MRCTに組み入れた時点で、「一定の均一性あり」と判断しているはずであり、情報借用を考えても良いのでは。
  - Phase II 試験データ等に基づき、情報借用の妥当性を検討すべきでは。
  - 全体平均に縮小するのではなく、国・地域をいくつかの集団にまとめる（パラメータを別々に設定する）ことも検討すべきでは。
    - 分け方については中間時点の成績に基づき考える。
    - 地理的・医療水準など、国・地域の分け方（クラスタリング）をいくつかのパターンでやってみて、ある種の頑健性を評価する。
  - 用いるモデルを工夫することも一案
    - 複数のモデルを仮定する。
    - 交換不可能な場合も考慮したモデルを用いる（EXNEXモデル等）

## 議論結果③

- その他

- 情報借用の程度が強すぎるのでは（例えば、Icelandの結果）

- イベント数が少ない国・地域における解析結果については解釈に注意が必要と考えられるため、イベント数が一定程度確保できるように試験デザインを工夫すべきでは。

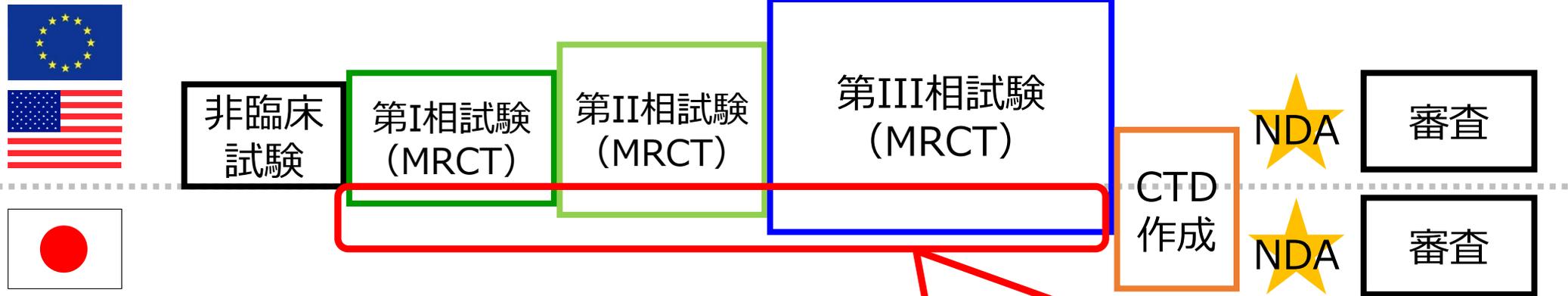
- 情報借用の程度を変えられるような工夫を考えてはどうか。

- イベント数も考慮して、国・地域のまとめ方（パラメータの設定）を考える等

- 情報借用の程度を少しずつ変えた解析（tipping point）を行ってはどうか。

議題2:  
第I～III相試験を通じた  
包括的な有効性評価

# “真の”世界同時開発に向けて



- 第III相試験における日本人症例数が非常に少なくなる
- 一方、承認申請には日本人の有効性・安全性評価が必要  
→Co-dataの枠組みを利用することで、第I相・第II相試験データを利用して精度高く日本人の有効性・安全性評価ができるのでは？

# Co-dataの利用例：第III相試験における無益性中間評価 (Neuenschwander et al. (2016))

- 循環器疾患を対象とした第III相試験
- 評価指標：全生存期間
- 試験薬・対照薬に1:1ランダム割り付け
- 過去試験：PoC試験、第II相試験→ヒストリカルデータとして利用
- 第III相試験を2本並行して実施→互いのデータを同時データとして利用
- 各試験、
  - 150イベント以上観察された時点で無益性に関する中間解析\*を1回実施
    - \*成功確率（379イベント観察された時点で統計的有意差を示す予測確率）に基づき判定
  - 379イベント観察された時点で最終解析（仮説検定）

# Co-dataの利用例：第III相試験における無益性中間評価 (Neuenschwander et al. (2016))

**Table 1.** Interim futility analyses for phase III trials A and B: estimates of hazard ratio (95% probability interval) and log-hazard-ratio (standard deviation), probability that hazard ratio is less than 1, probability of phase III success (PoS), and effective sample size (ESS). For trials A and B, the required final number of events is 379.

Study	Deaths	HR (95%-int)	log(HR) (sd)	pr(HR < 1)	PoS	ESS
Stratified analyses						
1. Proof-of-concept	8	0.70 (0.18,2.80)	-0.36 (0.71)	0.69		8
2. Phase II	85	0.75 (0.49,1.15)	-0.29 (0.22)	0.91		85
3. Phase III study A	162	0.83 (0.61,1.13)	-0.19 (0.16)	0.88	0.45	162
4. Phase III study B	150	0.78 (0.57,1.07)	-0.25 (0.16)	0.94	0.64	150
Co-data analysis						
3. Phase III study A	162	0.80 (0.63,1.04)	-0.22 (0.13)	0.95	0.51	254
4. Phase III study B	150	0.79 (0.61,1.01)	-0.24 (0.13)	0.97	0.65	252

ヒストリカル  
データ

同時データ

各試験データのみを  
使った場合

Co-dataを  
利用した場合

# Co-dataの利用例：第III相試験における無益性中間評価 (Neuenschwander et al. (2016))

**Table 1.** Interim futility analyses for phase III trials A and B: estimates of hazard ratio (95% probability interval) and log-hazard-ratio (standard deviation), probability that hazard ratio is less than 1, probability of phase III success (PoS), and effective sample size (ESS). For trials A and B, the required final number of events is 379.

Study	Deaths	HR (95%-int)	log(HR) (sd)	pr(HR < 1)	PoS	ESS
Stratified analyses						
1. Proof-of-concept	8	0.70 (0.18,2.80)	-0.36			8
2. Phase II	85	0.75 (0.49,1.15)	-0.29 (0.22)	0.7		85
3. Phase III		0.83 (0.61,1.13)	-0.19 (0.16)	0.88	0.45	162
4. Phase III		0.78 (0.57,1.07)	-0.25 (0.16)	0.94	0.64	150
Co-data analysis						
3. Phase III study A	162	0.80 (0.63,1.04)	-0.22 (0.13)	0.95	0.51	254
4. Phase III study B	150	0.79 (0.61,1.01)	-0.24 (0.13)	0.97	0.65	252

95%区間が狭くなった

どちらも大きくなった

利用するデータ量が増えた

# Co-dataの利用例：第III相試験における無益性中間評価 (Neuenschwander et al. (2016))

**Table 2.** Sensitivity analyses for probability of success (%) for trial A, trial B, and both trials, for three co-data models and three hazard ratio scenarios for trial A and B ( $HR_A$ ,  $HR_B$ ) at interim.

Interim hazard ratios for trial A and B: $HR_A = 0.83$ , $HR_B = 0.78$			
	trial A	trial B	trials A and B
Full exchangeability	51	65	36
Differential heterogeneity	49	64	34
Exchangeability-nonexchangeability mixture	49	65	34
Interim hazard ratios for trial A and B: $HR_A = 1.00$ , $HR_B = 0.78$			
	trial A	trial B	trials A and B
Full exchangeability	8	57	6
Differential heterogeneity	7	57	5
Exchangeability-nonexchangeability mixture	7	60	5
Interim hazard ratios for trial A and B: $HR_A = 0.50$ , $HR_B = 0.78$			
	trial A	trial B	trials A and B
Full exchangeability	100	78	78
Differential heterogeneity	100	76	76
Exchangeability-nonexchangeability mixture	100	72	72

- どちらも $HR < 1$ の場合（上）
  - Full exchangeabilityが最もtrial A/trial B/A and BのPoSが高い
- 1試験のみ有効性が見られなかった場合（中）
  - mixtureが最もtrial BのPoSが高い
- Trial Aの有効性が高かった場合（下）
  - Full exchangeabilityが最もtrial B/A and BのPoSが高い（trial Aの結果に引っ張られすぎ？）

# 論点

- 日本人について、第I～III相試験を通じた包括的な有効性評価を行う上で、気を付けることは何か？
- 第I～III相試験の情報を効率的に利用する方法はあるだろうか？

## 議論結果①

- 第Ⅲ相試験における第Ⅰ～Ⅱ相試験成績等の利用可能性について
  - 同じプロジェクトの第Ⅱ相試験は利用し易いのでは。
  - 第Ⅲ相試験の無益性中止に利用することは良いと考えるが、有効性中止に用いるのは難しいのではないか。
  - 第Ⅱ相試験がサロゲートエンドポイントに基づき評価されている場合は利用が難しいのではないか。
  - 第Ⅱ相試験の（日本人）症例数は少ないので、情報を借りる利点がないのでは。
  - 第Ⅱ相試験のデータが古いと問題になるのでは（レジメンの変更など）。
  - アウトカムのばらつきが非常に大きい疾患には適用しづらいだろう。

## 議論結果②

- 第Ⅲ相試験における第Ⅰ～Ⅱ相試験成績等の利用可能性について（続き）
  - 開発期間全体をとおして、プラセボデータを収集する。
  - Oncology領域では第Ⅰ～第Ⅲ相試験で試験デザインが大きく異なるので、前相の情報を利用するのは難しいかもしれない。
- 第Ⅲ相試験における第Ⅰ～Ⅱ相試験成績等の利用する際の留意点
  - 第Ⅰ相及び第Ⅱ相試験の結果は過大評価になっている可能性があるので、利用する場合は割り引かないといけないのでは。
    - 事前に解析を計画しておくことは重要と考える。
  - 試験間における集団の違いは考慮しなければならない。
  - エンドポイントの評価方法が異なる場合には、エンドポイントの評価に係る検査のタイミングや測定方法などに注意が必要と考える。

## 議論結果③

- 試験間での情報利用が期待される領域
  - オープン領域や再生医療分野
  - 再現性がある、かつ大規模な臨床試験は出来ない領域
    - 抗菌薬の小児とか
    - それぞれの試験単体では検出力不足だという状況での利用
    - 例えば2つの試験のうち結果の良くない方の情報量を増やすようなイメージの使い方はできるのではないか。
- Co-dataの動的利用について
  - シミュレーションで性能を十分に事前確認することが重要と考える。
    - どれくらいのシナリオを検討すればよいか、どういった性能であれば十分であるか（例えば $\alpha$ エラーのインフレ確認など）は今後の検討が必要であろう。
  - 動的利用で性能が劇的に変わるかはデータの量次第

# 参考文献

- Neuenschwander et al. On the Use of Co-Data in Clinical Trials. *Statistics in Biopharmaceutical Research* 2016; 8(3): 345-354.
- Lim et al. Minimizing Patient Burden Through the Use of Historical Subject-Level Data in Innovative Confirmatory Clinical Trials: Review of Methods and Opportunities. *Therapeutic Innovation & Regulatory Science* 2018; 52(5): 546-559.
- 武田他. 臨床試験におけるヒストリカルコントロールデータの利用. *計量生物学* 2015; 36(1): 25-50.
- Quan et al. Empirical shrinkage estimator for consistency assessment of treatment effects in multi-regional clinical trials. *Statistics in Medicine* 2013; 32(10):1691-1706.
- Guo et al. Evaluation of local treatment effect by borrowing information from similar countries in multi-regional clinical trials. *Statistics in Medicine* 2016; 35(5): 671-684.
- 厚生労働省医薬・生活衛生局医薬品審査管理課長. 国際共同治験の計画及びデザインに関する一般原則に関するガイドラインについて. 薬生薬審発0612第1号.  
<https://www.pmda.go.jp/files/000224557.pdf>
- 小宮山靖. E17に基づいてどのようなことが考えられ、我々に何をもたらすか？ What does ICH E17 imply and bring us? <https://www.mhlw.go.jp/content/11123000/000335160.pdf>