

実例にみる階層的な臨床的仮説が存在する場合 のがん臨床試験デザインについて

下川敏雄
和歌山県立医科大学

臨床試験とゲートキーピングへのニーズ

実例にみる階層的な臨床的仮説が存在する場合のがん臨床試験デザインについて
下川敏雄 (和歌山県立医科大学)

■ Key secondary endpointをおくこと

Primary
endpoint

not significant

secondary
endpoint

significant

通常の試験デザインの場合には、「副次的評価項目はデザインされていない」ということで、有意差が認められても多重性の問題から、あくまでも探索的な結果として扱われる。

とくに最近の臨床試験では、Key secondary endpointを設定したうえで、その必要症例数についても考慮したもとのデザインを行う場面が増えている。なぜなら、主要評価項目に有意差があった場合でも、その有効性の範囲(レスポンス、あるいは効果の範囲)についての詳細なプロファイリングが必要になっているためである。

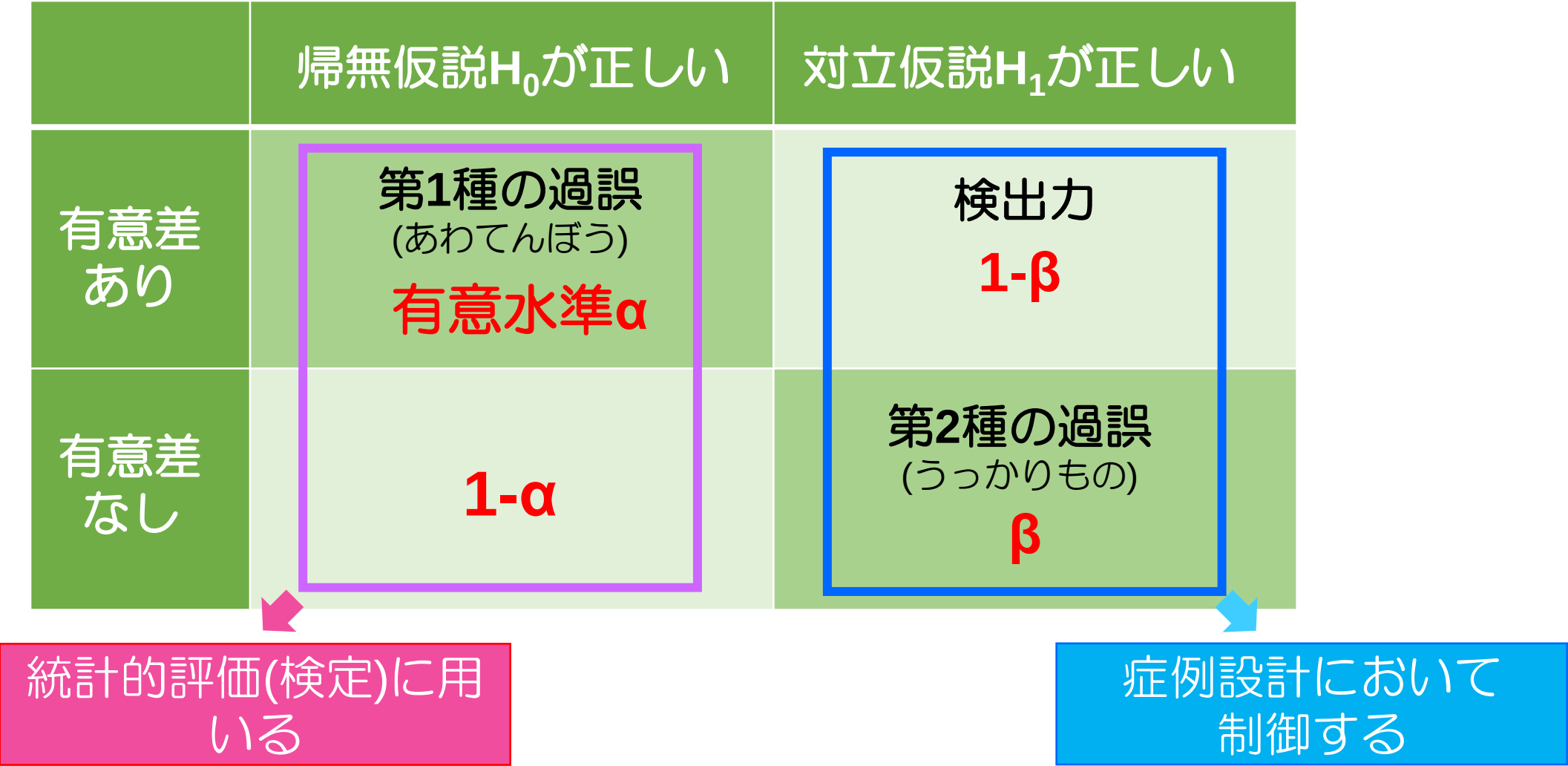
■ 複数の主要評価項目を立てる場合

- すべての主要評価項目が有意でないとPositive studyといえない場合
- いずれかの主要評価項目が有意であれば、Positive studyといえる場合

本日の話では、上記の場面のいくつかを想定したうえで、実際の臨床試験の実例とともに話します。

準備：第1種の過誤と第2種の過誤

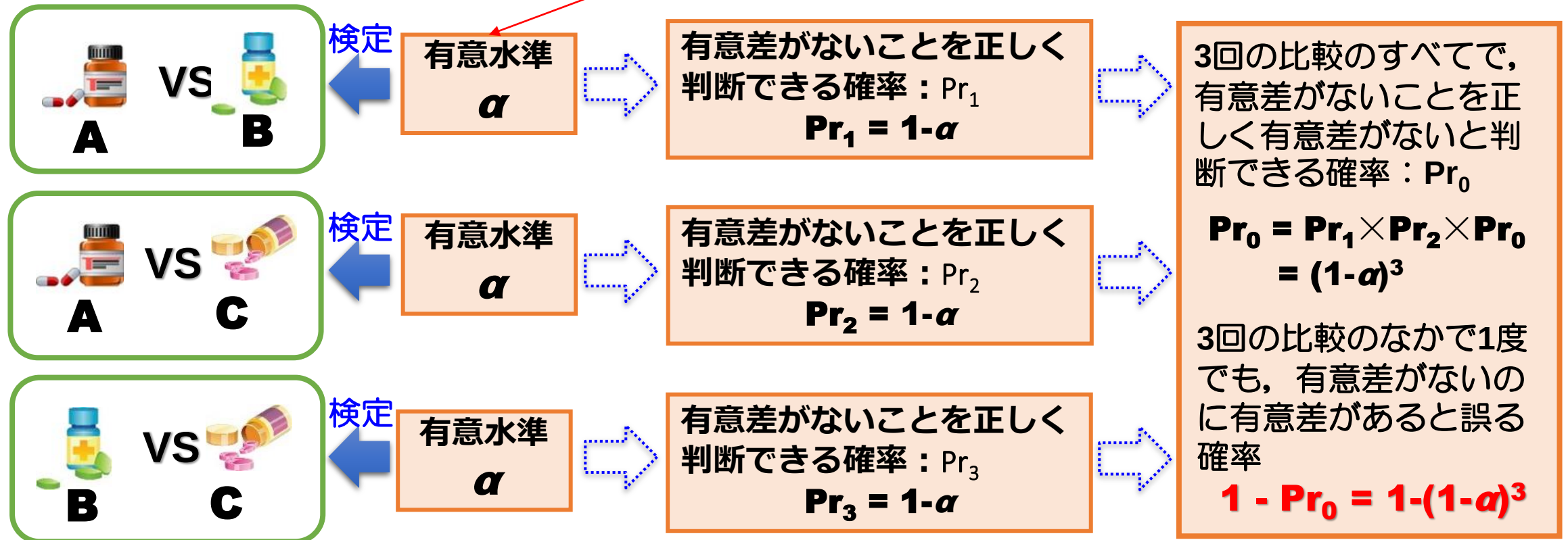
実例にみる階層的な臨床的仮説が存在する場合のがん臨床試験デザインについて
下川敏雄 (和歌山県立医科大学)



本日のお話しは、複数の評価項目に対して検定を繰り返すことによる多重性の問題(事前に設定した有意水準 α を上回ってしまう)を解決するための手順のゲートキーピング法である。

準備：多重比較はなぜ必要か？

有意水準とは「有意差がないのに有意差があると誤る確率(第1種の過誤)」をあらわす



1回の検定の有意水準 α を次のように考える：

$\alpha=0.05$ (1回の検定において有意差がないのに有意差があると誤る確率を0.05とする)

3回の検定を繰り返した場合、3回の検定の中で1度でも有意差がないのに有意差があると誤る確率は、

$$1 - (1 - \alpha)^3 = 1 - (1 - 0.05)^3 = 1 - (1 - 0.05)^3 = 0.142$$

となり、誤りの確率が増大してしまう。

もっと平たく言いますと. . .

いま、奏効率が5%の抗がん剤があったとする. この薬を1名の患者に投与した場合, その患者が奏効する確率は. . .

 5%

です(当たり前ですが. . .). 一方で, この抗がん剤を3名に投与したときに, 1名でも奏効すれば, この薬が有効であると判断するとします. 1名でも奏効すればよいということは, 少なくとも1名以上が奏効する確率ということになるので,

Pr = 1- (1名も奏効しない確率)

= 1- (個々の被験者で奏効しない確率)³

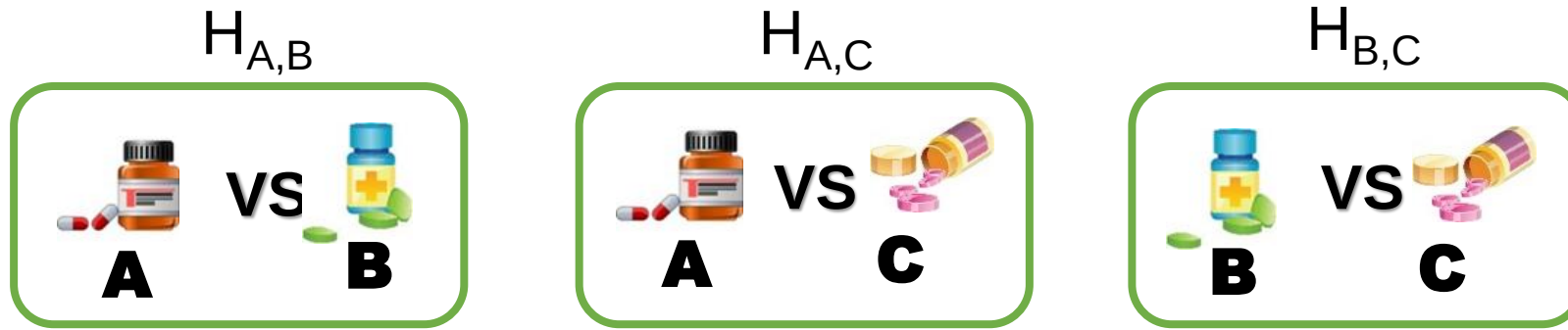
= 1-0.95³ = 0.142



になります(確率の簡単な問題です). つまり, 14.2%の確率で, この抗がん剤が有効であると判断されます. 言い方を変えれば, 「下手な鉄砲も数打てば当たる」ということになります. 検定も同じで, 第1種の過誤を0.05 (5%)としていても, 検定を繰り返すことは, 上のようなことをしているということになります. それを調整するために行う作業(procedure)が多重比較になります.

FamilyWise Error Rate (FWER)

多重比較では、複数の帰無仮説 H_0 を検定する。先ほどの3種類の薬剤の比較の場合には、



の3個の帰無仮説を検定する。このとき、 $\mathcal{F} = \{H_{A,B}, H_{A,C}, H_{B,C}\}$ をファミリー(family)という。

このファミリーにおいて、正しい帰無仮説のうち、少なくとも1つを棄却してしまう確率をFWER (familywise error rate)という。多重比較とは、このFWERを制御することを意味する。

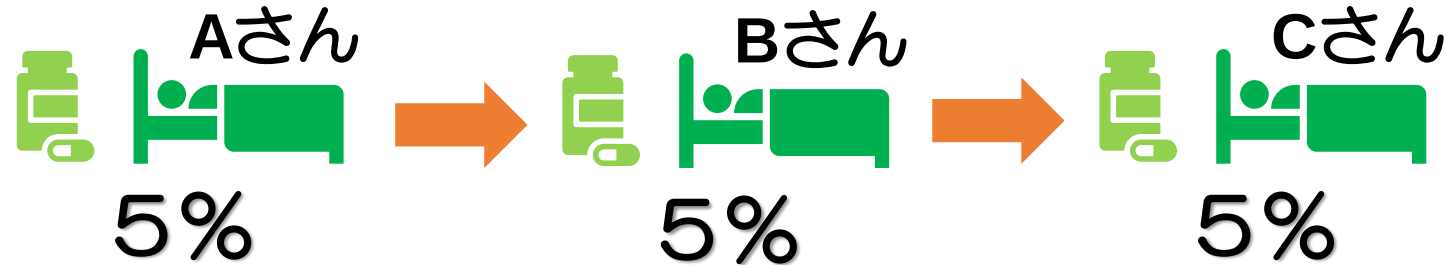
階層仮説(主にGatekeeping)においても、FWERの制御が目的である。つまり、階層的な試験デザインでは、「いずれかの帰無仮説が正しい場合においてもFWERを有意水準 α 以下に保つ」ようにデザインされる。

仮説を階層化すること：ゲートキーピング法の目的

いま、奏効率が5%の抗がん剤があったとする。この薬を1名の患者に投与した場合、その患者が奏効する確率は. . .



です(当たり前ですが. . .)。一方で、この抗がん剤を3名(A,B,C)に投与したとします。このとき、Aさんが奏効したらBさんに投与し、Bさんが奏効したらCさんに投与する、と順番をつける とします。



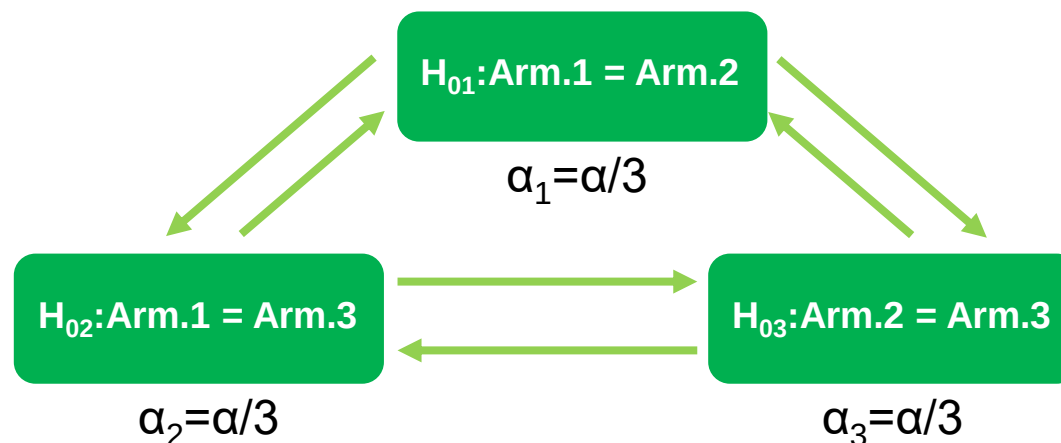
Aさんが奏効した後にBさんを評価したとき、Bさんの期待される奏効率に変化はありません(Aさんが奏効したからと言ってBさんの奏効率が変化する確率に変化は生じません)。つまり、シーケンシャルに仮説をおいてあげれば、FWERは、一定に保つことができます。この事実を複雑にしたり、していったものが、ゲートキーピング法を見ることができます。

多重比較と仮説の階層化について

臨床的重要度に順序関係が存在しない多重比較

Bonfferoni法

いま、ある3種類のレジメンに対する無作為化比較試験を検討している。Bonfferoni法による多重比較を行うと以下のようなになる。



Bonfferoni法による多重比較のステップは、下記のとおり：

Step.1 :

H_{01}

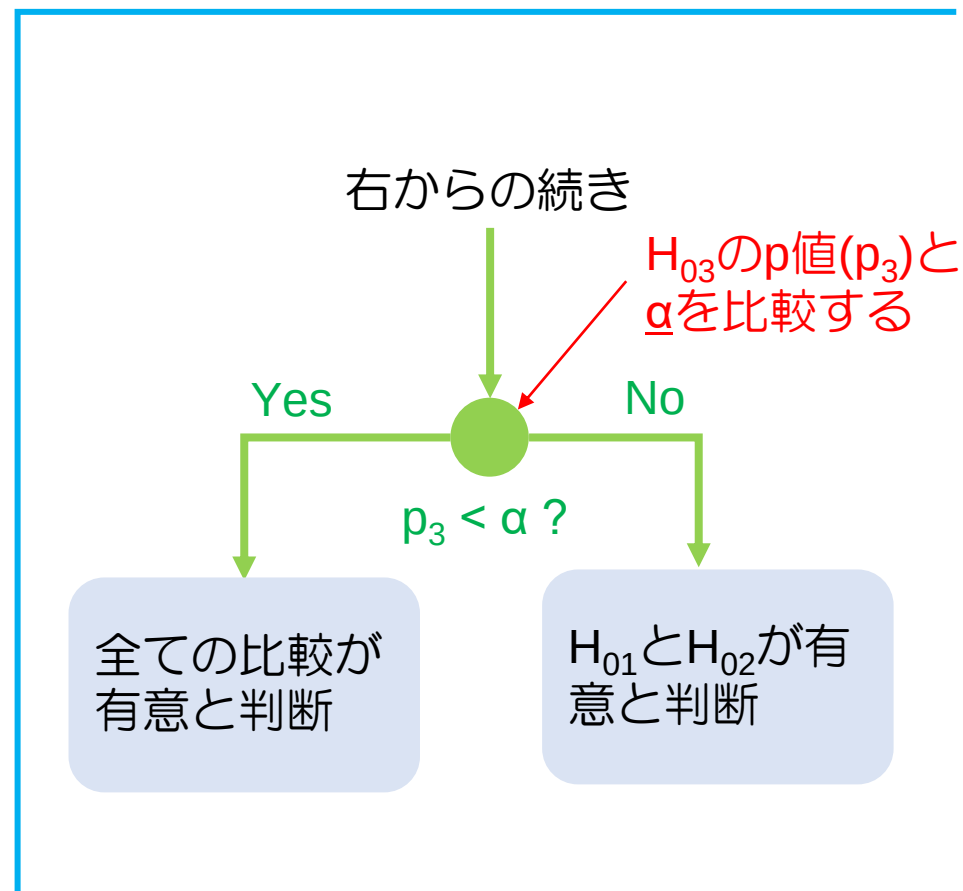
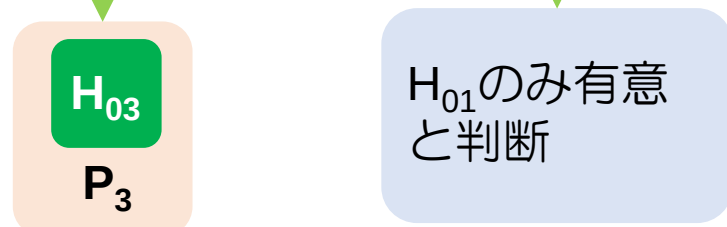
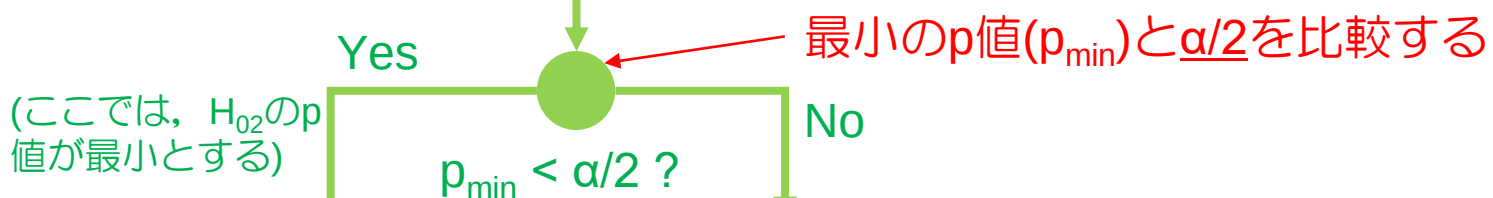
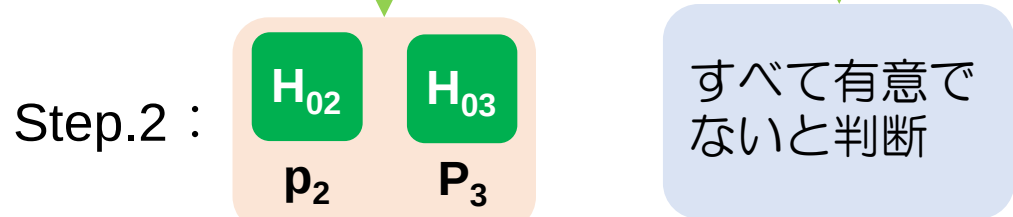
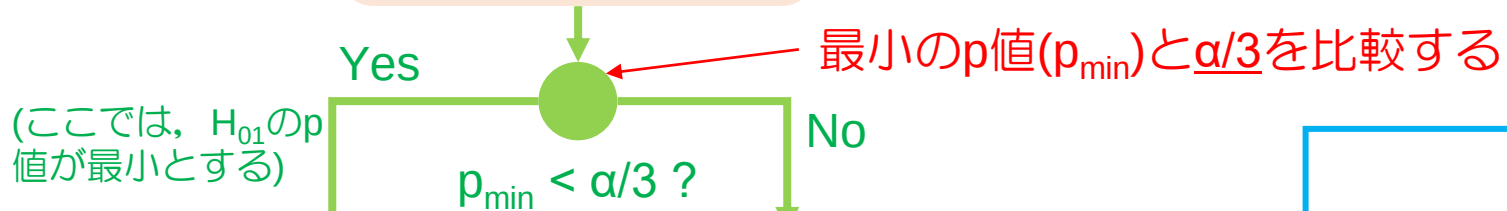
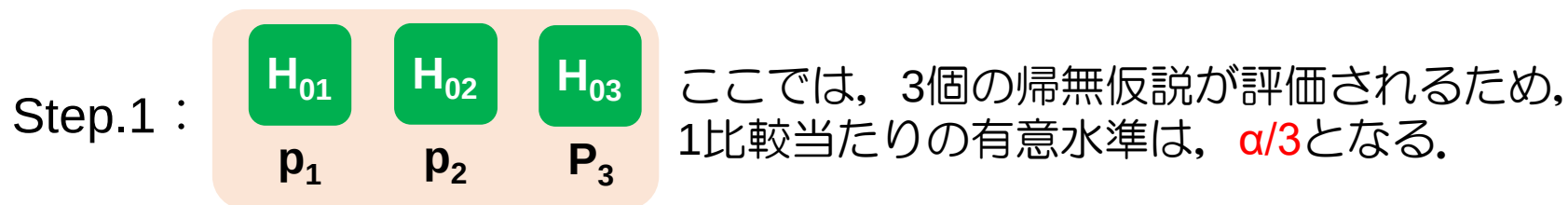
H_{02}

H_{03}

帰無仮説が棄却された比較に対して、有意差があるとみることができる。

Holm法

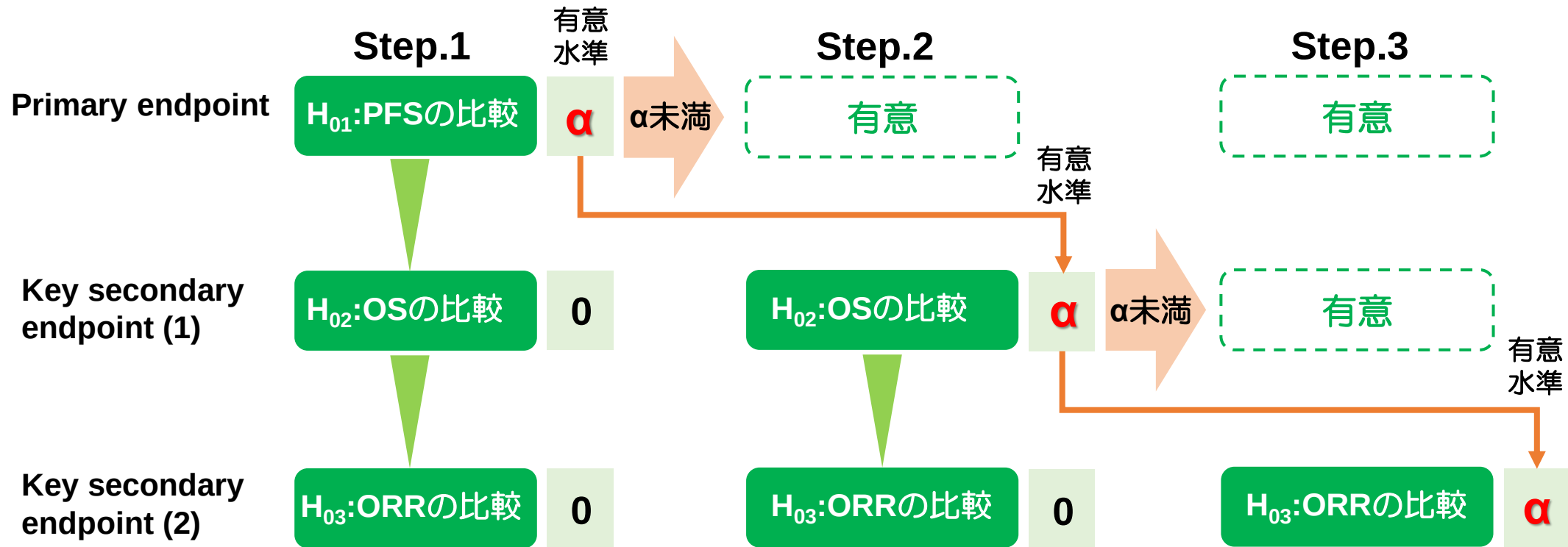
ステップダウン形式で多重比較を行うが、仮説間に臨床的な順序関係は存在しない。



臨床的重要度に順序関係が存在する多重比較

階層型手順(固定順序手順) Maurer et al. Biometrie in der Chemisch-Pharmazeutischen Industrie, 1995; 6:3-18.

いま、ある臨床試験を計画している。この試験では、Primary endpointにPFS、Key secondary endpointにOS、ORRを計画している。また、OSのほうがORRよりも重要視しているとする。

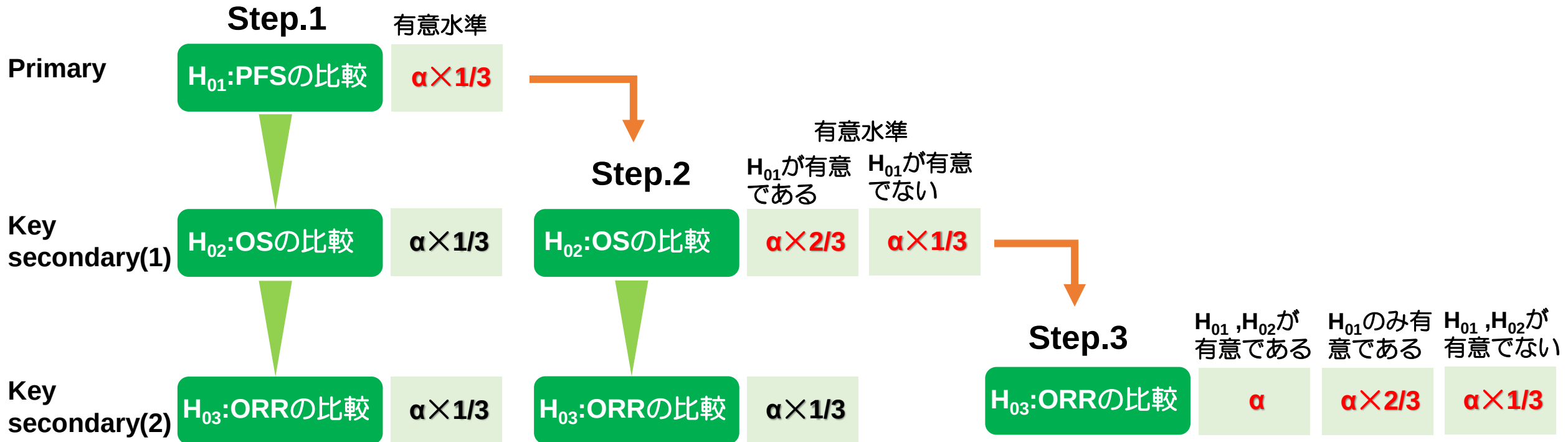


階層型手順では、1つのステップに対して1個の検定しかしないため、 **α のまま検定しても、FWERを α 以下に抑えることができる**。一方で、途中のステップで有意でない場合、以降のエンドポイントが有意であっても、参考資料にしかない(Bonferroni法, Holm法では有意であると認識できる)。

Fallback法

Maurer et al. Biometrie in der Chemisch-Pharmazeutischen Industrie, 1995; 6:3-18.

いま, ある臨床試験を計画している. この試験では, Primary endpointにPFS, Key secondary endpointにOS, ORRを計画している. また, OSのほうがORRよりも重要視しているとする. ここでは, 例として, 各比較に $\alpha/3$ を割振ったとする(重みを変更することはできる).



Fallback法は, 階層型手順とHolm手順の中間的な方法である. この手順は, 階層型手順と異なり, ステップの途中で有意差がなくても, アルゴリズムを終了することなく, 評価を行えることである. 一方で, ステップの初期の有意水準は, Holm法と同じになるので, 有意差が出にくくなる(あるいは, デザイン段階での必要症例数が多く必要になる).

階層型手順のもとでの試験例

実例にみる階層的な臨床的仮説が存在する場合のがん臨床試験デザインについて
下川敏雄 (和歌山県立医科大学)

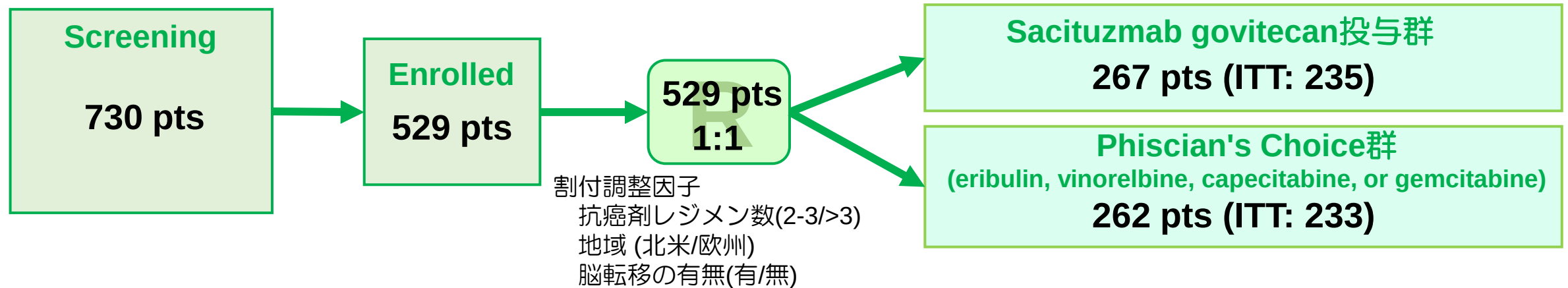
ORIGINAL ARTICLE

Sacituzumab Govitecan in Metastatic Triple-Negative Breast Cancer

Bardia et al. NEJM, 2021; 384:1529-1541.

転移性トリプルネガティブ乳癌に対するsacituzumab
govitecan療法の国際共同無作為化比較第III相試験

Flow of Patients



Outcome

Primary endpoint：無増悪生存期間,

Key Secondary endpoints：全生存期間

Secondary endpoints：奏効割合, 奏効までの期間, 奏効期間, 安全性

階層型手順のもとでの試験例

実例にみる階層的な臨床的仮説が存在する場合のがん臨床試験デザインについて
下川敏雄 (和歌山県立医科大学)

Family.1
($\alpha=0.05$)

PFS
($\alpha=0.05$)

PFSに関して、 $HR=0.667$ (Active MST=4.5, Control MST=3.0), **有意水準 $\alpha=0.05$ (2-sided)**, 検出力 $1-\beta \geq 0.95$ と設定した場合に、脳転移のない患者に対して315例のイベントが必要になる (**予定登録数: 488例**: うち脳転移患者に対して15%のキャップをつける).

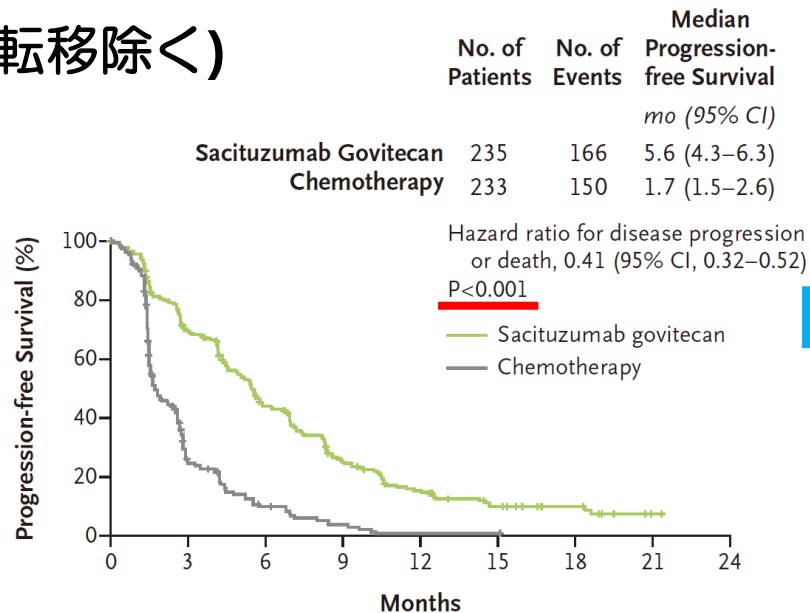
Significant

Family.2
($\alpha=0.05$)

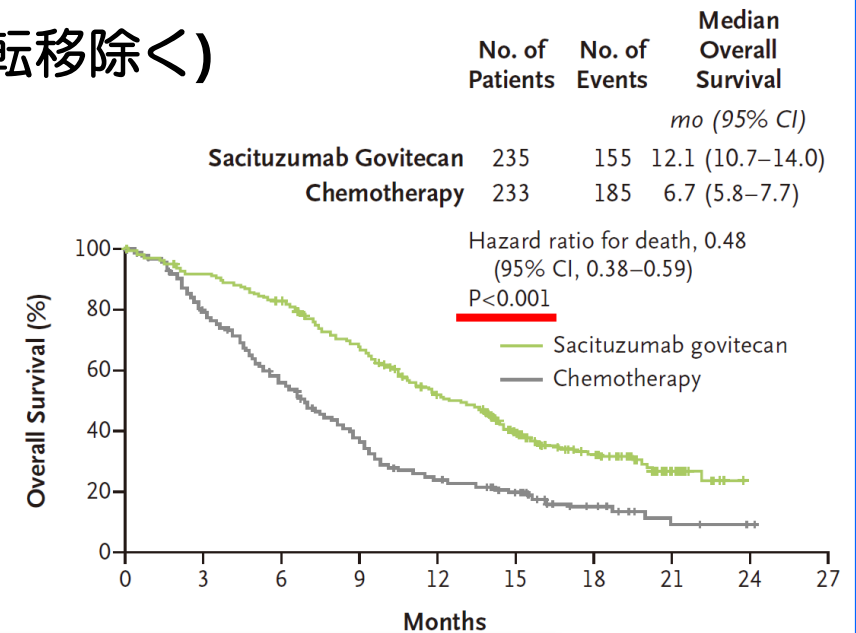
OS
($\alpha=0.05$)

OSに関して、 $HR=0.70$ (Active MST=14.3, Control MST=10.0), **有意水準 $\alpha=0.05$ (2-sided)**, 検出力 $1-\beta \geq 0.89.5$ と設定した場合に、脳転移のない患者に対して330例のイベントが必要になる. OSの解析は、予定症例数330例(脳転移のない患者)に到達した時点で実施される.

PFS (脳転移除く)



OS (脳転移除く)

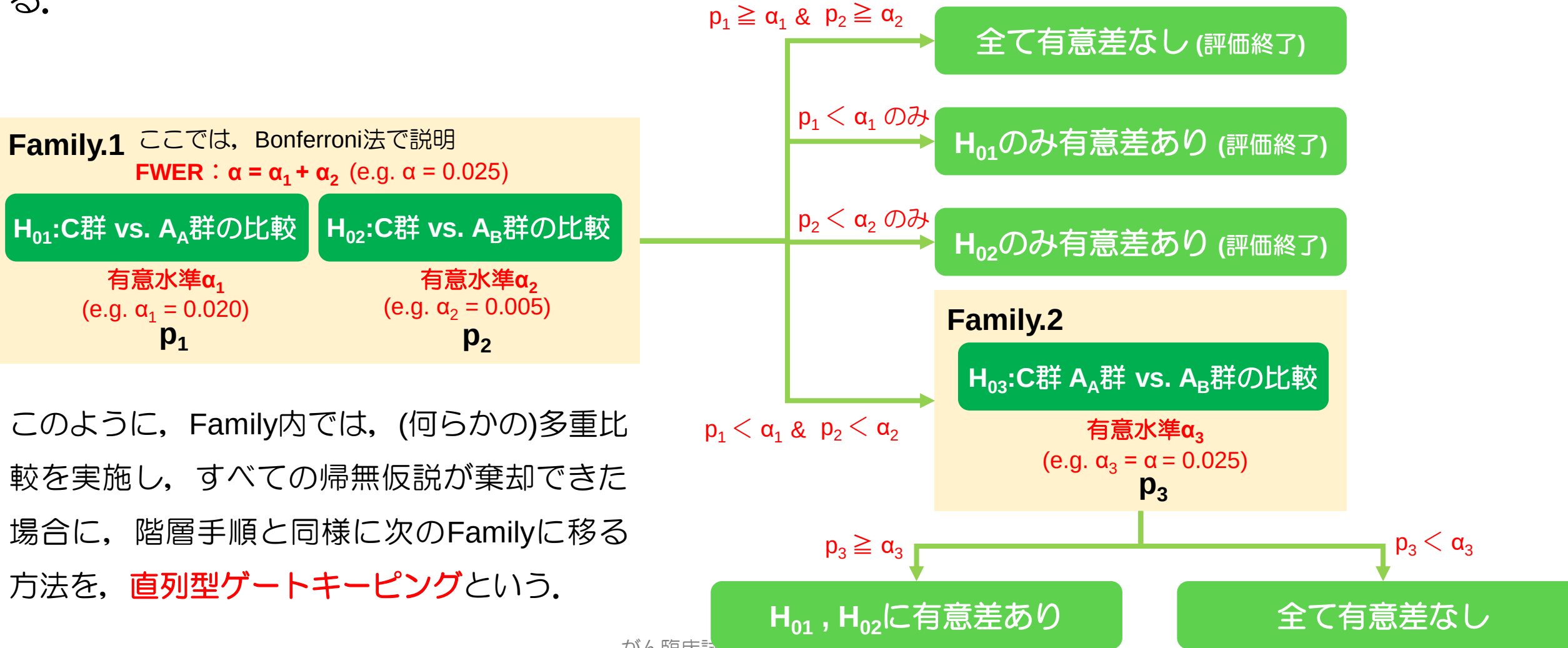


PFS, OSともに, Sacitumab Govitecanの有効性を検証することができた.

37	13	0
11	3	1

直列型ゲートキーピング法

階層手順では、Familyが一つの仮説のみで構成されていた、一方で次のような状況も考えられる。
いま、対照群(C群)に対して試験薬A群(A_A 群)および試験薬B群(A_B 群)が存在する。本試験での主要評価項目は、C群 vs. A_A 群、C群 vs. A_B 群 OSであり、Key secondary endpointが A_A vs. A_B のOSであるとする。



このように、Family内では、(何らかの)多重比較を実施し、すべての帰無仮説が棄却できた場合に、階層手順と同様に次のFamilyに移る方法を、**直列型ゲートキーピング**という。

直列型ゲートキーピング法を利用した試験例

実例にみる階層的な臨床的仮説が存在する場合のがん臨床試験デザインについて
下川敏雄 (和歌山県立医科大学)

ORIGINAL ARTICLE

Encorafenib, Binimetinib, and Cetuximab
in *BRAF* V600E-Mutated Colorectal Cancer

Kopetz et al. NEJM, 2019; 381:1632-1643.

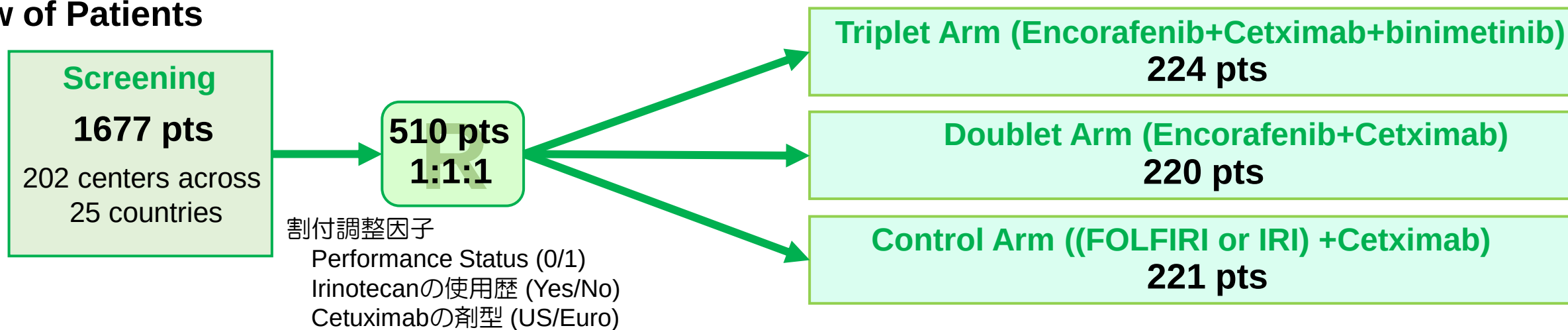
**Encorafenib Plus Cetuximab as a New Standard
of Care for Previously Treated *BRAF* V600E-
Mutant Metastatic Colorectal Cancer: Updated
Survival Results and Subgroup Analyses from the
BEACON Study**

Tabernero et al. JCO, 2021; 39(4):273-285.

BRAF V600E変異大腸癌に対する

**Encorafenib, Binimetinib, Cetuximab
の併用療法に関する国際共同無作為化
比較第III相試験 (BEACON study)**

Flow of Patients



Outcome

Primary endpoint : 全生存期間 (Triplet vs. Control), ORR(Triplet vs. Control)[後ほど追加]

Key Secondary endpoints : 全生存期間(Doublet vs. Control)

Secondary endpoints : ORR, PFS, TTR(Time to response), DOR(Duration of Response), PK, QOLの各群組み合わせ

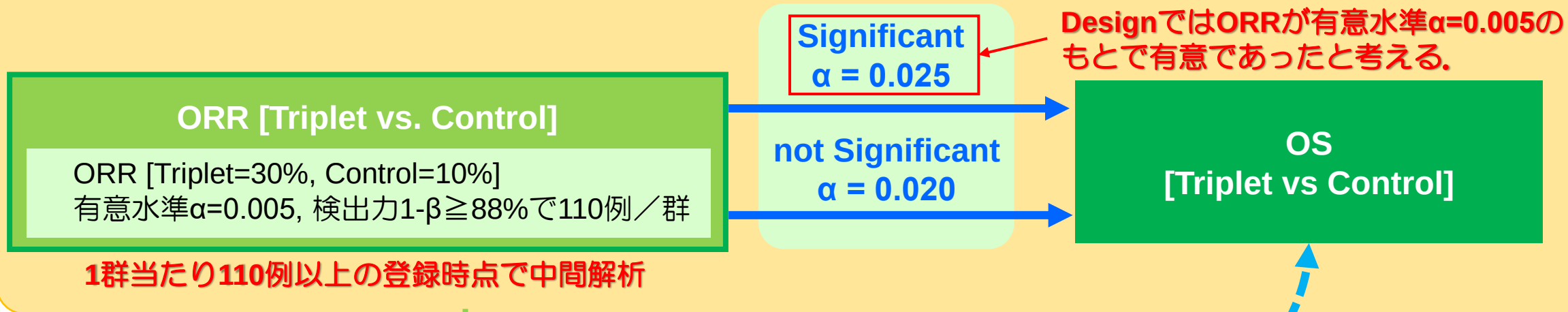
直列型ゲートキープングを利用した試験例

実例にみる階層的な臨床的仮説が存在する場合のがん臨床試験デザインについて
下川敏雄 (和歌山県立医科大学)

Family. 1

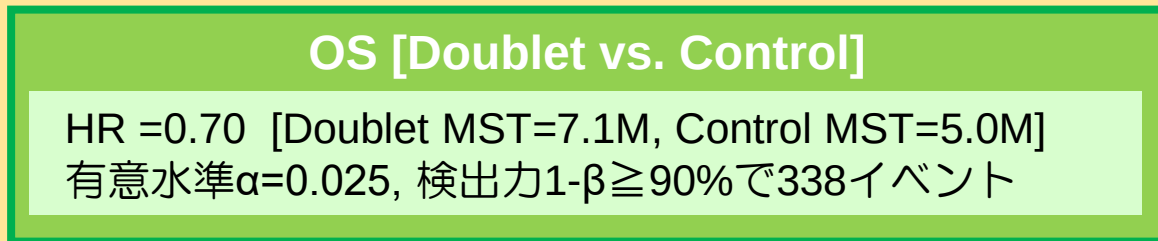
Primary endpoints

このように、一方の帰無仮説が棄却された場合にそのときの有意水準 α を引き継ぐ形での多重比較をfallback解析という。



Family. 2

Key secondary endpoint



上記のイベント数を試験期間内にカウントするために,215例/群と設定している。

Tripletのほうが, DoubletよりもOSが延長することを仮定しているため(Triplet MST=7.5M, Doublet MST=7.1M), Doublet vs. Controlが有意であれば, Triplet vs. Controlも有意である。
なお, 上記の設定において, Triplet+Controlのイベント数は268になるが, HR=0.67, 有意水準 $\alpha=0.025$ での検出力は, 約90%になる。

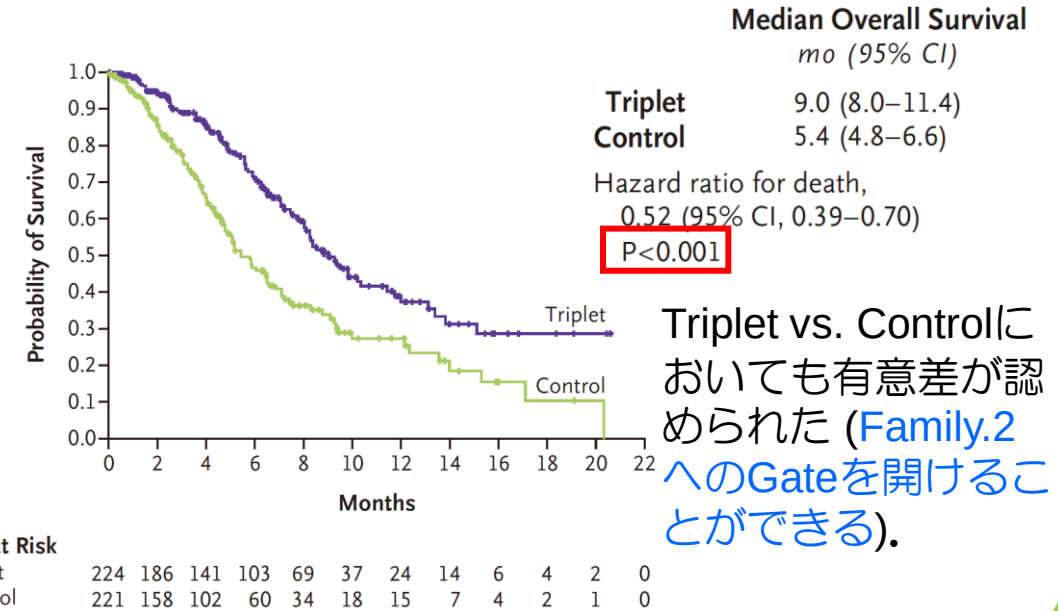
Note: ここでの有意水準 α は片側(1-sided)である。

Family.1 Primary endpoints

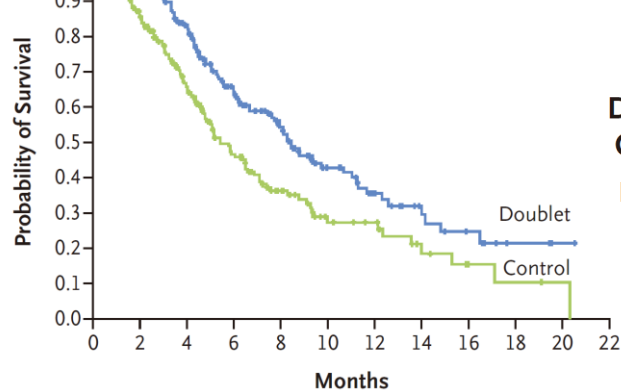
	Triplet Regimen (N = 111)	Doublet Regimen (N = 113)	Control (N = 107)
CR, PR (%)	29 (26)	23 (20)	2 (2)
95%CI	18–35	13–29	<1–7
p-value (vs. control)	<0.001	<0.001	

0.005未満なので、 $\alpha=0.025$ (1-side)のもとでOSを評価

Fallback



Family.2 Key secondary endpoint



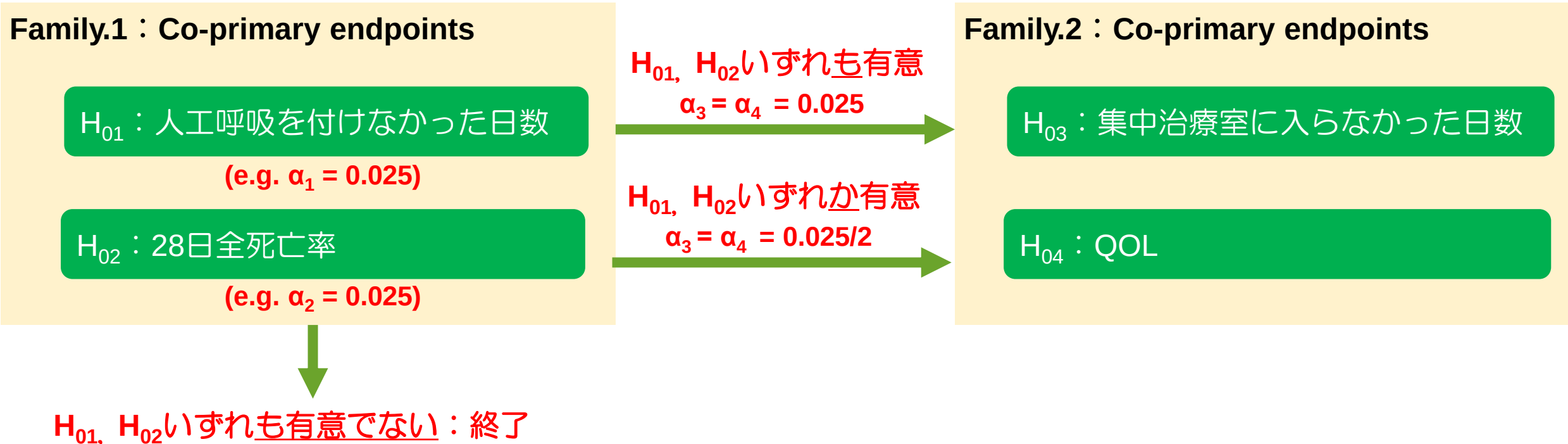
したがって、Triplet, Doubletともに、Controlに比べて、OS, ORRともに優越性が検証できた。

なお、本試験は、中間解析時点で有効性が認められたため、その時点で試験結果がNEJMで公表されている。最終報告は、Subgroup解析を含めてJCOで公表されている。

並列型ゲートキーピング法

直列型ゲートキーピング法では、Family内のすべての帰無仮説をすべて棄却した場合のみ、次のステップに移行することができた。これに対して、Co-primary endpointsのいずれかを満たせば次のステップに移行する場合も存在する。このような場合に用いられるのが、並列型ゲートキーピング法である。

Dmitrienko et al. (2013)は、急性呼吸不全を対象としたプラセボ対照試験を例にして、並列型ゲートキーピング法を説明している。



並列型ゲートキーピング法の例示

JAMA Oncology | Original Investigation

**Durvalumab With or Without Tremelimumab vs Standard Chemotherapy
in First-line Treatment of Metastatic Non-Small Cell Lung Cancer**
The MYSTIC Phase 3 Randomized Clinical Trial

Naiyer et al. JAMA Oncol., 2020; 6(5):661-674.

実例にみる階層的な臨床的仮説が存在する場合のがん臨床試験デザインについて
下川敏雄 (和歌山県立医科大学)

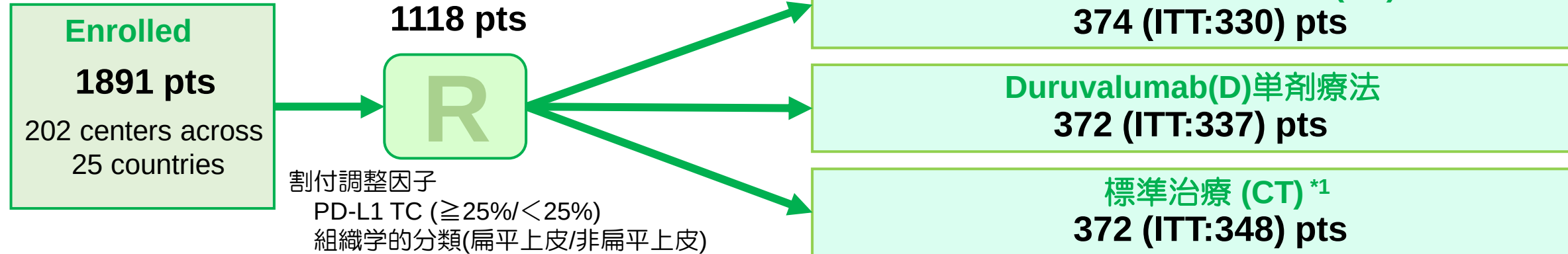
非小細胞肺癌に対する1次治療としての

Durvalumab + Tremelimumab併用,

Durvalumab単剤, 標準治療に対する国際共同無

作為化第III相試験 (MYSTIC trial)

Flow of Patients



*1: プラチナ製剤+ α のDoublet chemotherapy

Outcome

Primary endpoint : PD-L1 TC $\geq 25\%$ 症例でのOS (DT vs. CT, D vs. CT), PFS(DT vs. CT)

Secondary endpoints : PFS, OS(PD-L1 TC $\geq 25\%$, PD-L1 TC $\geq 21\%$, ITT)において, Primary endpoint以外
DT vs. CTおよびD vs. CTでのORR, 治療成功期間, 忍容性, 安全性

並列型ゲートキーピング法の例示：必要症例数の計算

Primary endpoints

$\alpha=0.005$

H_{01} : DT vs. CT (PFS)

PD-L1 TC $\geq 25\%$

HR = 0.59 (1yr. survival: DT=34.1, CT=13.0%)

power = 88% (イベント割合72%)

231 events

$\alpha=0.030$

H_{02} : D vs. CT (OS)

PD-L1 TC $\geq 25\%$

HR = 0.62 (18M survival: D=53.2, CT=36.0%)

power = 90% (イベント割合70%)

225 events

$\alpha=0.015$

H_{03} : DT vs. CT (OS)

PD-L1 TC $\geq 25\%$

HR = 0.62 (18M survival: D=53.2, CT=36.0%)

power = 86% (イベント割合70%)

225 events

■ H_{01} における必要症例数： $231 \times 1/0.72 \doteq 320$ 例

■ H_{02}, H_{03} における必要症例数： $225 \times 1/0.70 = 320$ 例

3群で考えれば， $160 \times 3 = 480$ 例 (160pts/arm)

480例 (160pts/arm)は，あくまでもPD-L1 TC $\geq 25\%$ の症例である．そのため，想定されるITT症例に占めるPD-L1 TC $\geq 25\%$ を見積もる必要がある．プロトコール内には明記されていないが，1092例(364pts/arm)あれば，予定症例数に達すると計画している (試験自体は，event drivenで進められる)．

並列型ゲートキーピング法の例示：ゲートキーピングの手順

Primary endpoints

$\alpha=0.005$

H_{01} : DT vs. CT (PFS)

PD-L1 TC $\geq 25\%$

HR = 0.59 (1yr. survival: DT=34.1, CT=13.0%)
power = 88% (イベント割合72%)

231 events

Significant

$\alpha=0.030$

H_{02} : D vs. CT (OS)

PD-L1 TC $\geq 25\%$

HR = 0.62 (18M survival: D=53.2, CT=36.0%)
power = 90% (イベント割合70%)

225 events

Significant

$\alpha=0.015$

H_{03} : DT vs. CT (OS)

PD-L1 TC $\geq 25\%$

HR = 0.62 (18M survival: D=53.2, CT=36.0%)
power = 86% (イベント割合70%)

225 events

Significant

Secondary endpoints

H_{04} : D vs. CT (PFS)

PD-L1 TC $\geq 25\%$

Significant

$\alpha=0.005$

H_{06} : DT vs. CT (PFS)

PD-L1 TC $\geq 1\%$

Significant

$\alpha=0.005$

H_{08} : DT vs. CT (PFS)

ITT

$\alpha=0.005$

H_{05} : DT vs. CT (OS)

PD-L1 TC $\geq 1\%$

Fallback procedure

- H_{02} : significant, H_{03} : not significant $\rightarrow \alpha = 0.030$
- H_{02} : not significant, H_{03} : significant $\rightarrow \alpha = 0.015$
- H_{02} : significant, H_{03} : significant $\rightarrow \alpha = 0.030+0.015 = 0.045$

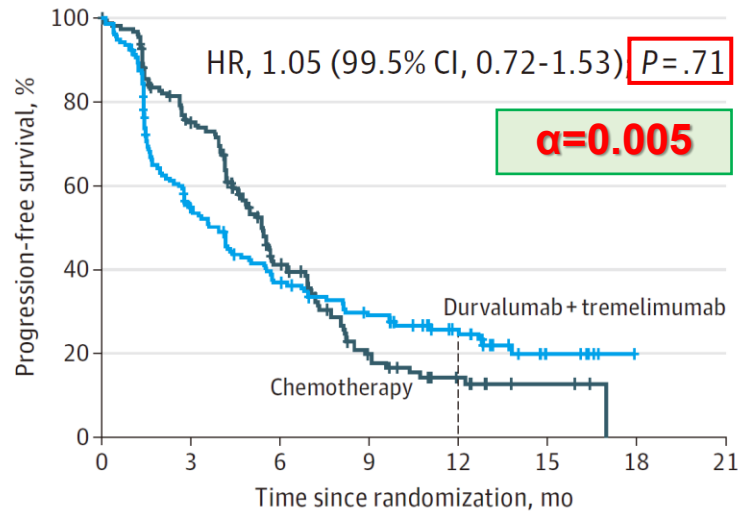
Significant

H_{07} : DT vs. CT (OS)

ITT

H_{05} の有意水準

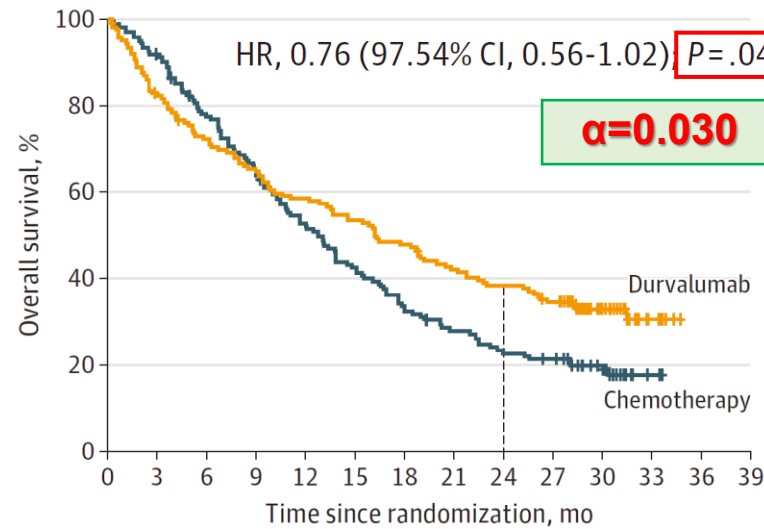
H₀₁ : DT vs. CT (PFS)
PD-L1 TC $\geq 25\%$



Not significant

H₀₄ : D vs. CT (PFS)
PD-L1 TC $\geq 25\%$

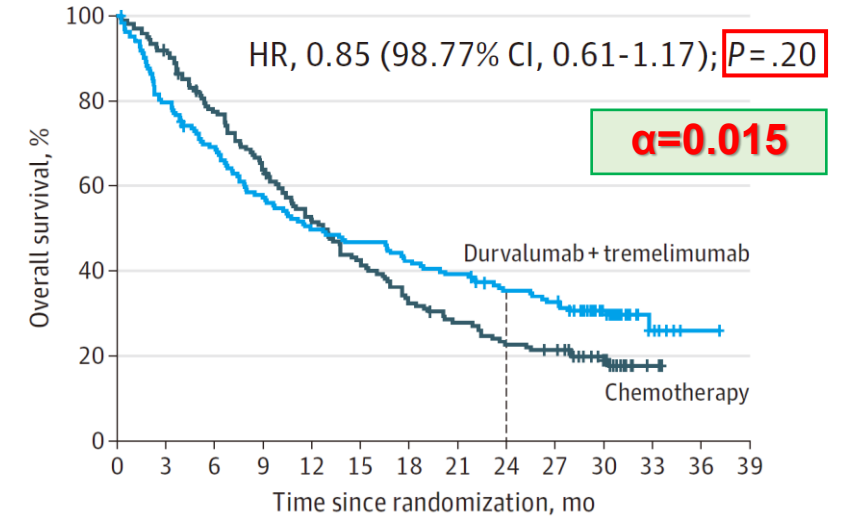
H₀₂ : D vs. CT (OS)
PD-L1 TC $\geq 25\%$



Not significant

H₀₅ : DT vs. CT (OS)
PD-L1 TC $\geq 1\%$

H₀₃ : DT vs. CT (OS)
PD-L1 TC $\geq 25\%$



Not significant

結局, すべてのPrimary endpointがmetしなかった. そのため, PD-L1 TCのその後の解析結果は, 論文では言及されていない(参考資料として, H₀₄が記載). 結局は, 血液中の遺伝子変異量(TMB; Tumor Mutational Burden)によるサブグループ解析のなかで, TMB ≥ 20 mut/Mbのサブグループにおいて, 顕著な差異が認められたことを述べている.

階層型構造のFamily内で多重比較を行う場合

先ほどの、急性呼吸不全を対象としたプラセボ対照試験の例を再検討する。並列型ゲートキーピング法では、Family1をBonferroni型の多重比較で行い、その結果に応じてFamily2に割振る有意水準 α が決まっていた。このとき、例えば、Family2をHolm法で構築することができる。

Bonferroni type procedure

Family.1 : Co-primary endpoints

H_{01} : 人工呼吸を付けなかった日数

(e.g. $\alpha_1 = 0.025$)

H_{02} : 28日全死亡率

(e.g. $\alpha_2 = 0.025$)

H_{01}, H_{02} いずれも有意

$\alpha_3 = \alpha_4 = 0.025$

H_{01}, H_{02} いずれか有意

$\alpha_3 = \alpha_4 = 0.025/2$

H_{01}, H_{02} いずれも有意でない : 終了

Holm type procedure

Family.2 : key secondary endpoints

H_{03} : 集中治療室に入らなかった日数

H_{04} が有意 : $\alpha_3 + \alpha_4$ H_{03} が有意 : $\alpha_3 + \alpha_4$

H_{04} : QOL

階層型構造のFamily内で多重比較を行った試験例

JAMA | Original Investigation

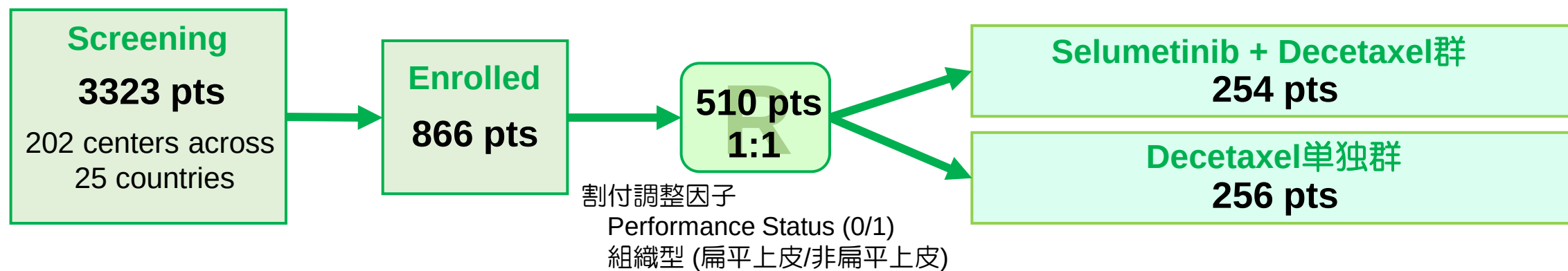
Selumetinib Plus Docetaxel Compared With Docetaxel Alone
and Progression-Free Survival in Patients With *KRAS*-Mutant
Advanced Non-Small Cell Lung Cancer
The SELECT-1 Randomized Clinical Trial

Jänne et al. JAMA, 2017; 317(18):1844-1853.

KRAS遺伝子変異を有する進行性非小細胞肺癌に対
するSelumetinib+Docetaxel併用療法 vs Docetaxel
単剤療法の国際共同無作為化比較第III相試験

(SELECT-1 study)

Flow of Patients



Outcome

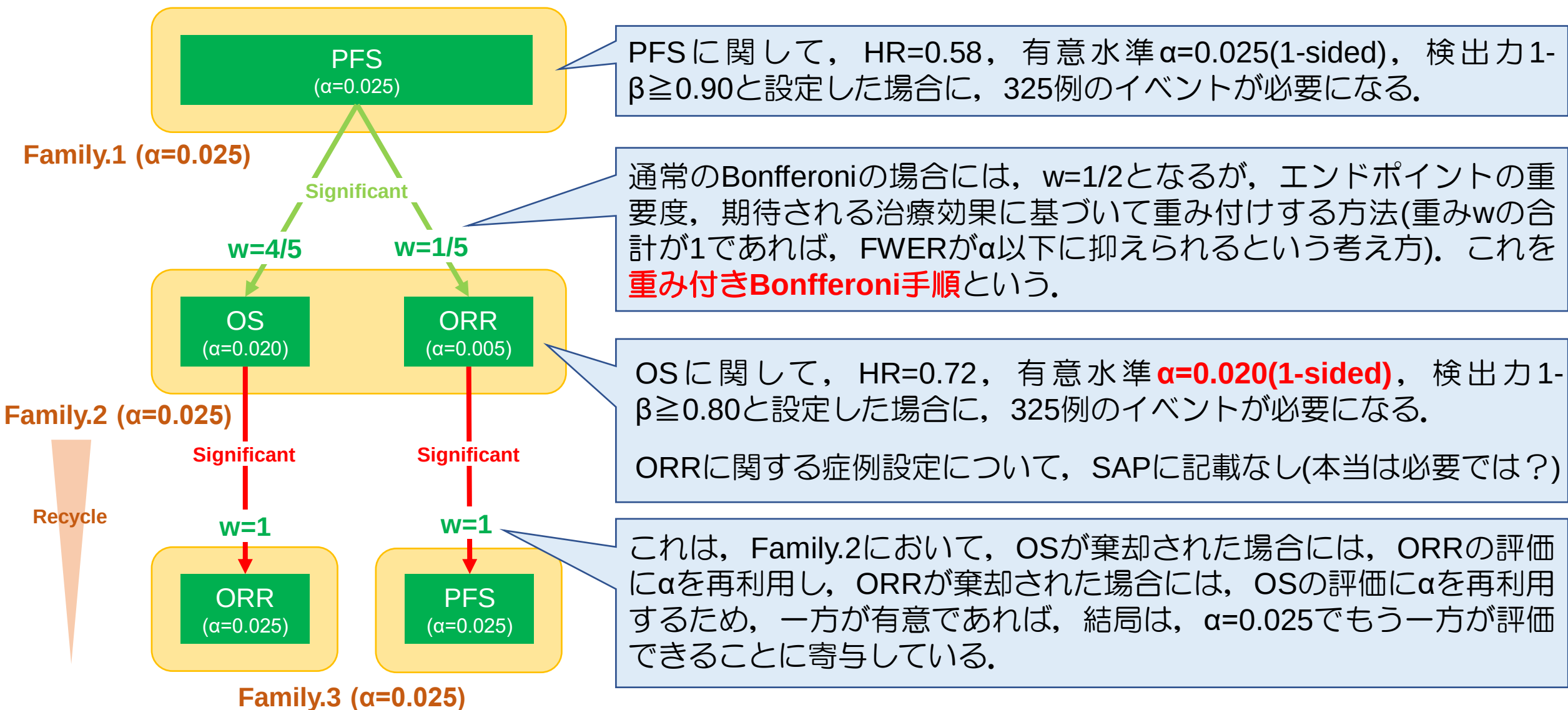
Primary endpoint：無増悪生存期間，

Key Secondary endpoints：全生存期間，奏効割合

Secondary endpoints：奏効期間，疾患関連症状への効果，安全性，忍容性

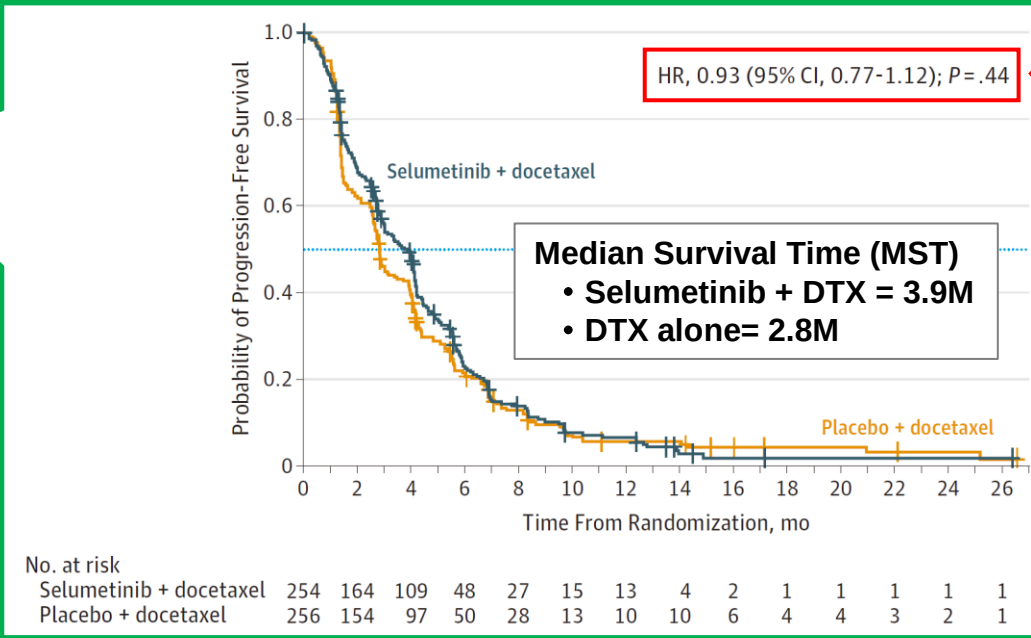
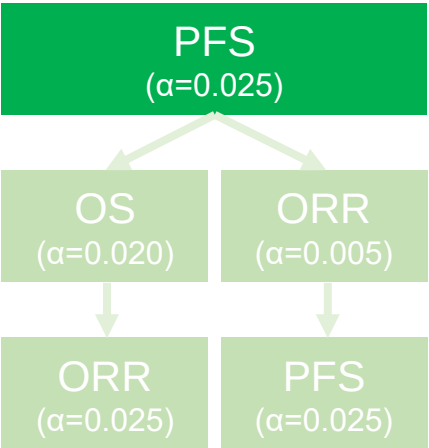
階層型構造のFamily内で多重比較を行った試験例

この試験では、**α再利用法(alpha-exhaustive recycling strategy, Burman et al., 2009)**という方法を利用している。ここでは、各Familyにおいて、割り当てられたαを再利用(recycle)するため、このような名前がついている。



Note：ここでの有意水準 α は片側(1-sided)である。

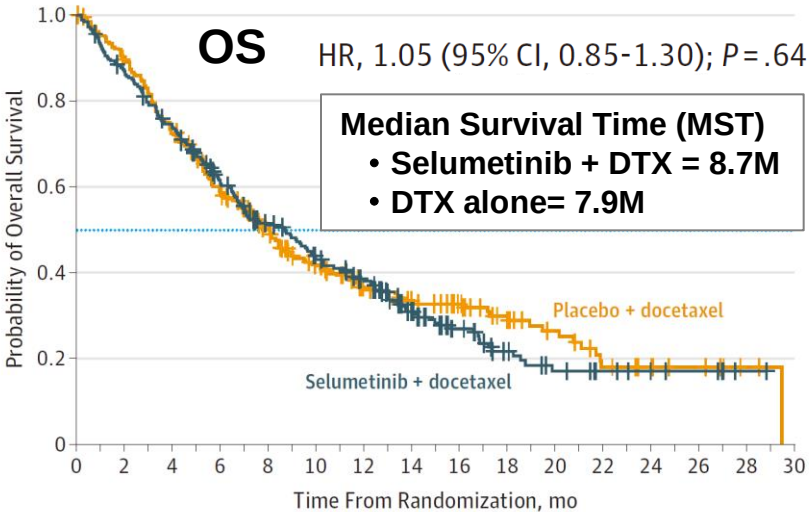
階層型構造のFamily内で多重比較を行った試験例



有意水準0.05のもとで有意差なし

後続のOS, ORRの検定は検証できない

仮にOS, ORRにおいて有意水準0.05のもとで有意差が認められたとしても、通常のPIIIのSecondary endpointの評価と同様の解釈。



	CR	PR	ORR	OR [95%CIs]
Selumetinib + DTX	2	49	20.1%	1.61 [1.00, 2.62]
DTX alone	0	35	13.7%	

p-value = 0.05

PFSが有意でない時点で試験としてはNegative studyになる。また、参考資料としてのOS, ORRともに有意ではなかった。Discussionでは、EGFRのサブタイプに言及しているが、エンドポイントに対するDiscussionは殆どないと思われる(大事なものは、きちんとデザインすれば、Negative studyでもトップジャーナルに載るということかも)。

より複雑な多重比較のための方法：グラフィカル・アプローチ

Bretz et al (2009)は、ゲートキーピングをグラフィカルに表す方法として、グラフィカル・アプローチを提案している。

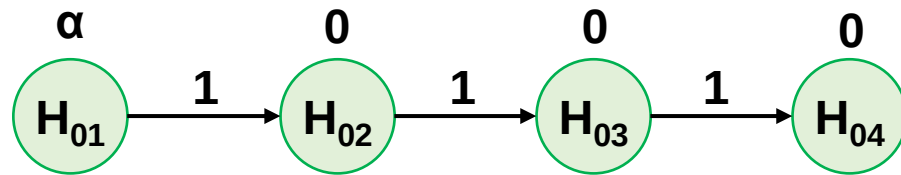
Bonferroni手順



○ は帰無仮説を表す(ノードと呼ばれる).
ノードのうえの $\alpha/4$ は、各帰無仮説に割振られた有意水準である(合計が α になるように割振る).

Bonferroni手順は、それぞれの帰無仮説を個別に取り扱うため、それぞれのノードに重み1/4で有意水準 α が割振られる(重みは合計が1になれば1/4以外でも良い(e.g: それぞれの有意水準の重みを1/8, 1/8, 1/2, 1/4にする)).

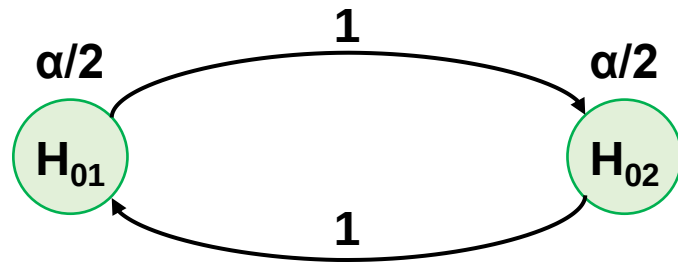
階層型手順



帰無仮説のあいだにシーケンスが存在する場合には、矢印で表し、そのうえに重みを割振る。

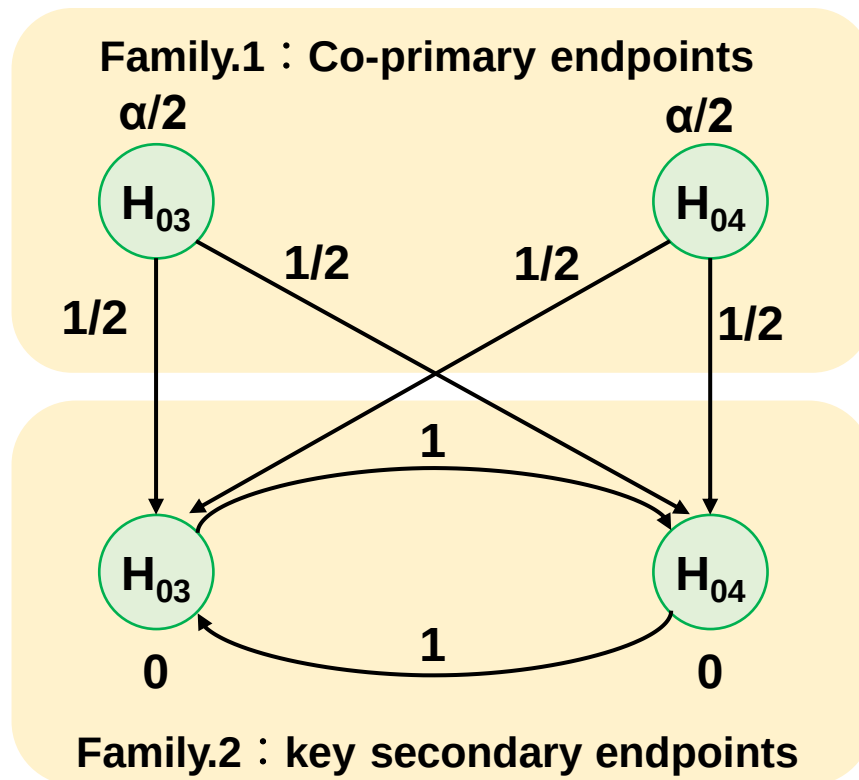
階層型手順では、 H_{01} から出発し、有意であれば、 H_{02} に移行する、このとき、有意水準 α を再利用するので、重みは1である。以降、同じ

Holm手順



Holm法では、一方の帰無仮説が棄却されれば(このとき、p値が小さい方を選ぶ)、もう一方の帰無仮説の有意水準 α_{sig} は、有意になった帰無仮説の有意水準 α に加えることができる。すなわち、双方向の矢印になる。

並列型ゲートキープング法



Bonferroni type procedure

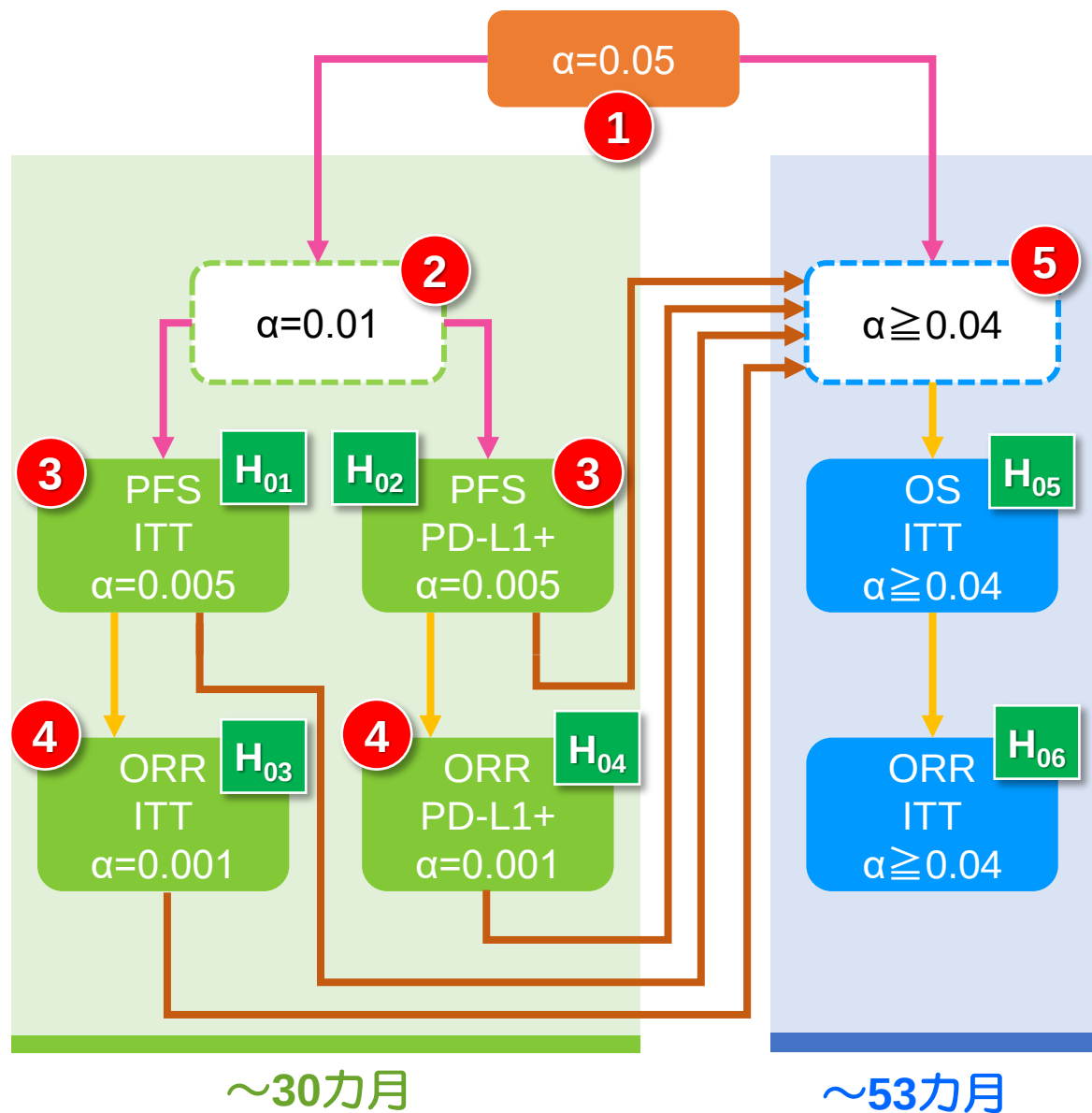
Holm type procedure

幾つかの多重比較を組み合わせた例示

ORIGINAL ARTICLE

Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer

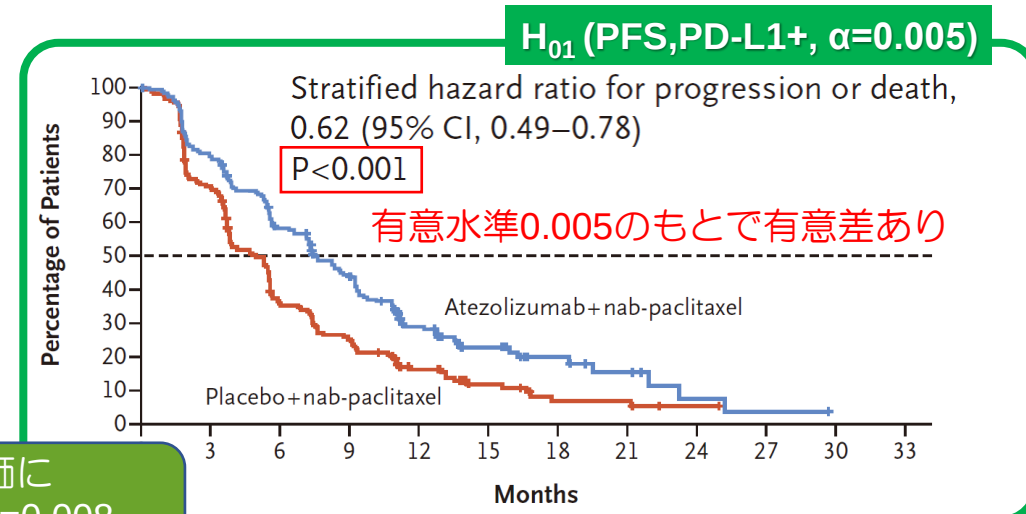
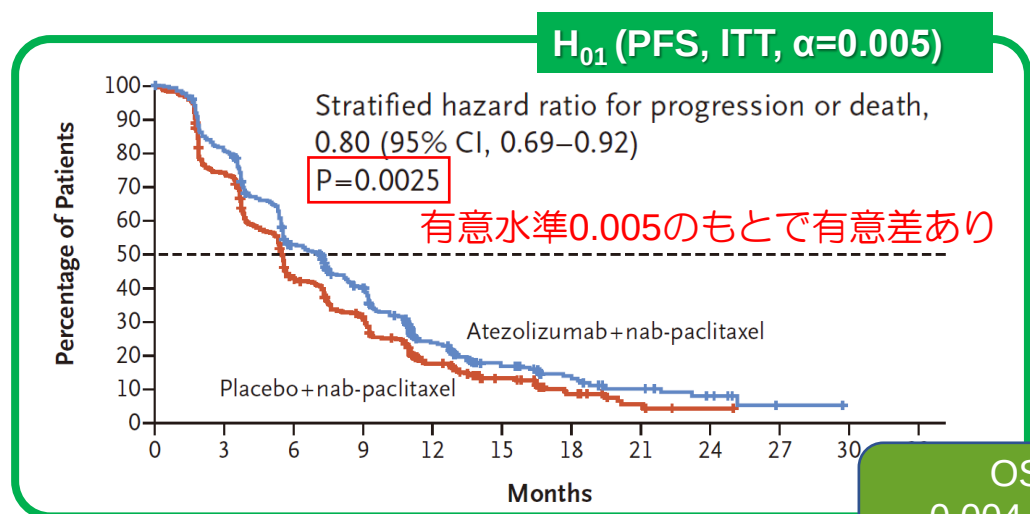
Schmid et al. NEJM, 2018; 379:2108-2121.



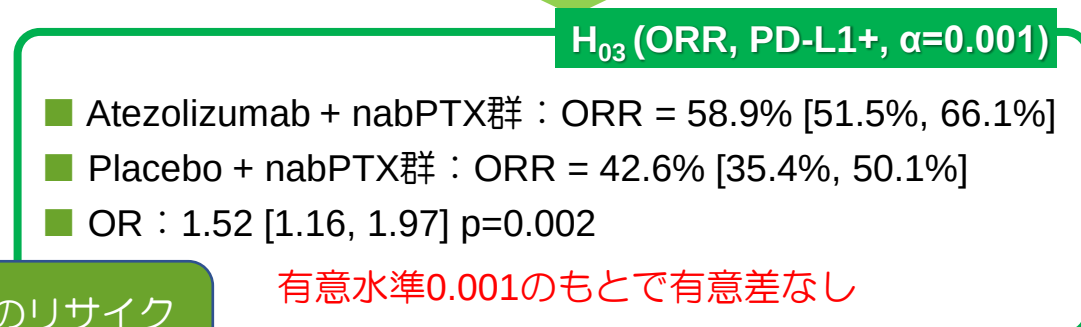
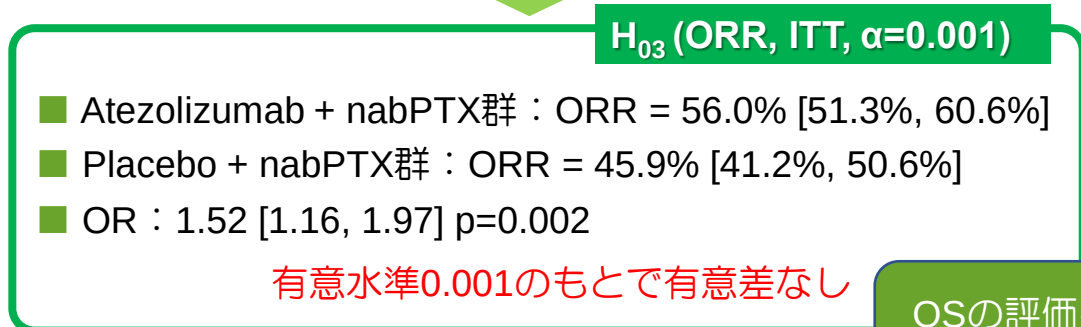
- ① OSとPFSのそれぞれの仮説に対して、有意水準 α を0.01と0.04に分ける(重み付きBonferroni)
- ② PFSの評価： $\alpha = 0.01$
 - ITT症例とPD-L1のサブグループのそれぞれに、 $\alpha = 0.005$ ずつ分ける(Bonferroni).
- ③ H_{01} , H_{02} のいずれか有意な場合には、階層型手順のもとで、次のステップに移るが、このとき、もとの $\alpha = 0.005$ のうち、0.001を H_{03} , H_{04} に振り、0.004をOSの評価にリサイクルする。
- ④ H_{03} , H_{04} のいずれか有意な場合には、その有意水準0.001をOSの評価にリサイクルする。
- ⑤ つまり、OSの評価における α は、0.04~0.05までの範囲をとる。

ゲートキーピング手順にあわせて解釈する

実例にみる階層的な臨床的仮説が存在する場合のがん臨床試験デザインについて
下川敏雄 (和歌山県立医科大学)



OSの評価に
 $0.004+0.004=0.008$
をリサイクル



OSの評価へのリサイク
ルなし

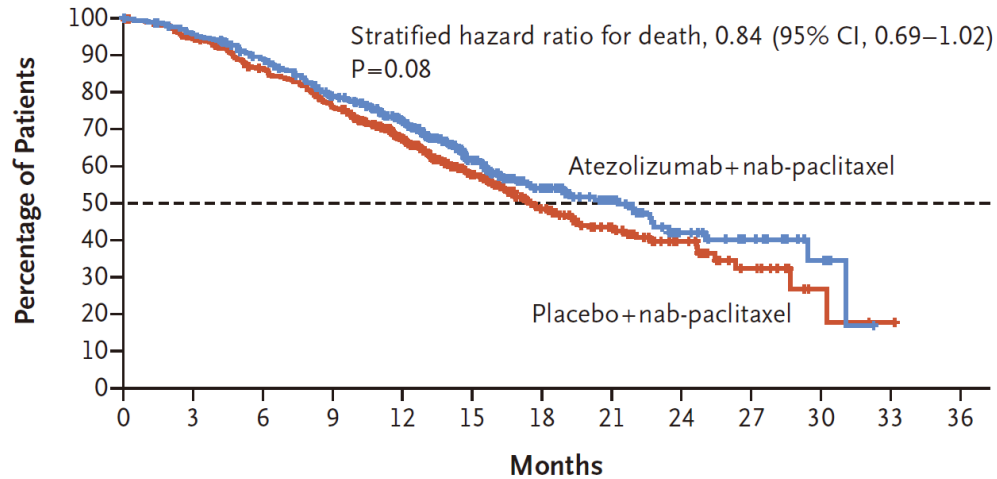
p-value = 0.05

OSの評価は, $0.04+0.008 = 0.048$ で行う.

ゲートキーピング手順にあわせて解釈する

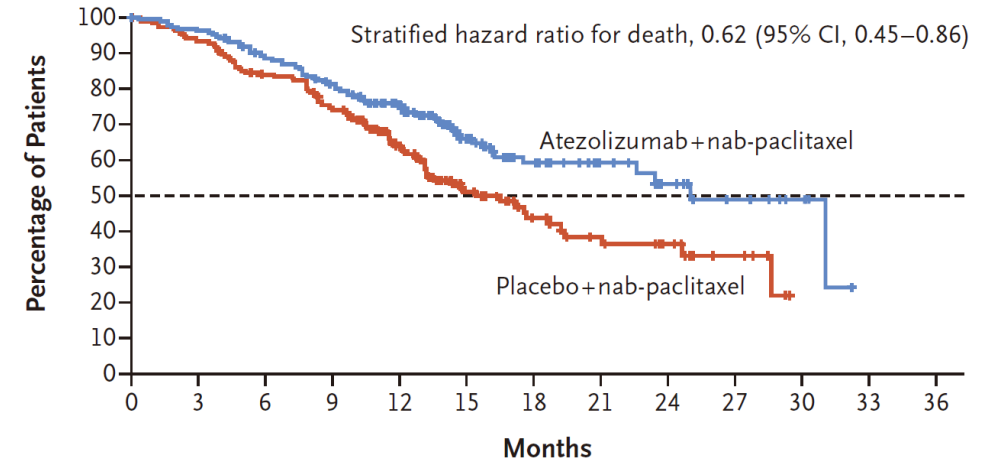
実例にみる階層的な臨床的仮説が存在する場合のがん臨床試験デザインについて
下川敏雄 (和歌山県立医科大学)

H_{05} (OS, ITT, $\alpha=0.048$)



p値は0.08だった。有意水準 $\alpha=0.048$ であることから、OSについては、Positiveではなかった。階層型仮説なので、ここでゲートキーピング手順は終了する。

H_{06} (OS, PD-L1+, α =なし(探索的))



H_{05} において、ゲートキーピングは終了しているので、p値が記載されていない。ただし、HRの95%信頼区間が1.0をまたいでいないので、有意水準 $\alpha=0.05$ のもとで有意であったと思われる(あくまでも参考)。

- ゲートキーピングに準じて解釈すると、PFSはITT症例、PD-L1+症例ともにPositiveである。
- 一方で、ORRおよびOSはPositiveであることが「検証」できなかった。
- ただし、ORR、OSともに有意水準 $\alpha=0.05$ でみれば有意であった(これまでと同様のSecondaryの評価でみれば、「検証はできなかったが違いはあった」程度のニュアンス)。

- 今回は、階層型臨床的仮説に対する方法として、ゲートキーピング法を紹介した。
- ゲートキーピング法が急激に用いられるようになった背景には、2017年に米国FDAのパブリックコメントドラフト「Draft Guidance for Multiple Endpoints in Clinical Trials」のなかで、FWERの取り扱いについて規制当局としての考え方が示された。
- 従来の臨床試験では、主要評価項目のみFWERを制御すればよかったのに対して、主要評価項目＋副次的評価項目の制御に考え方がシフトした。
- 今回紹介した国際共同無作為化比較第III相試験の殆どが、本邦での「治験」であり、積極的にゲートキーピング法を取り入れている。
- 一方で、論文では、論文中にゲートキーピングの手順について記載されることは殆どなく、試験デザインの多くがsupplementaryに記載されている。
- さらに、論文の書き方は従来と同じなので、わかりにくい（階層仮説では、有意水準が検定によって異なる場合がある）。p値が0.05未満でもnot significantと書かれている場合があるが、なぜ、有意でないのかについては、ゲートキーピング手順を知らないといけなくなる。今回の内容が、論文解釈の一助になることを祈念します。



Thank you for your kind attention

shimokaw@wakayama-med.ac.jp



toshibow2000@gmail.com

ここでは、ゲートキーピング手順の類似(FWERを有意水準 α 未満にするという意味では同じ)した試験デザインとして、バイオマーカーを伴う試験デザインについて記載する。

Enrichment Design



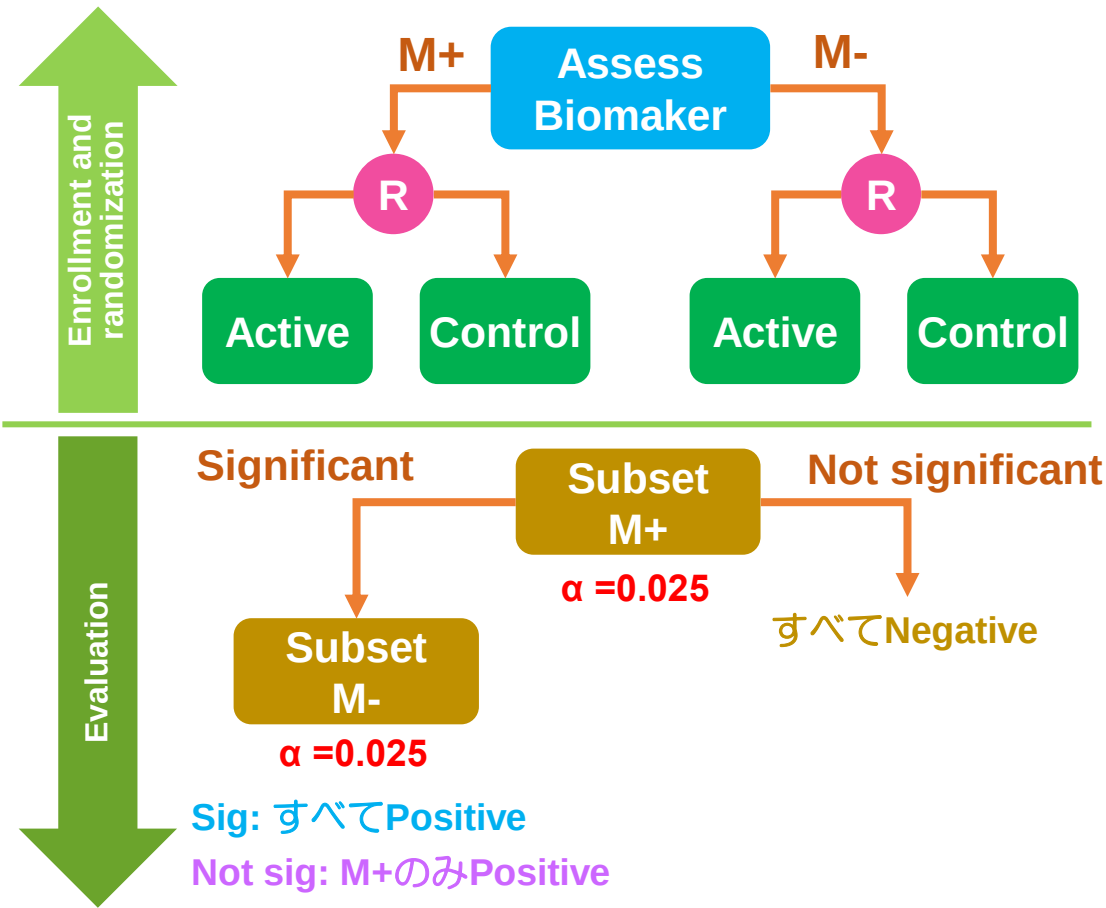
Subgroup Specific with recycling design



バイオマーカーを伴う試験

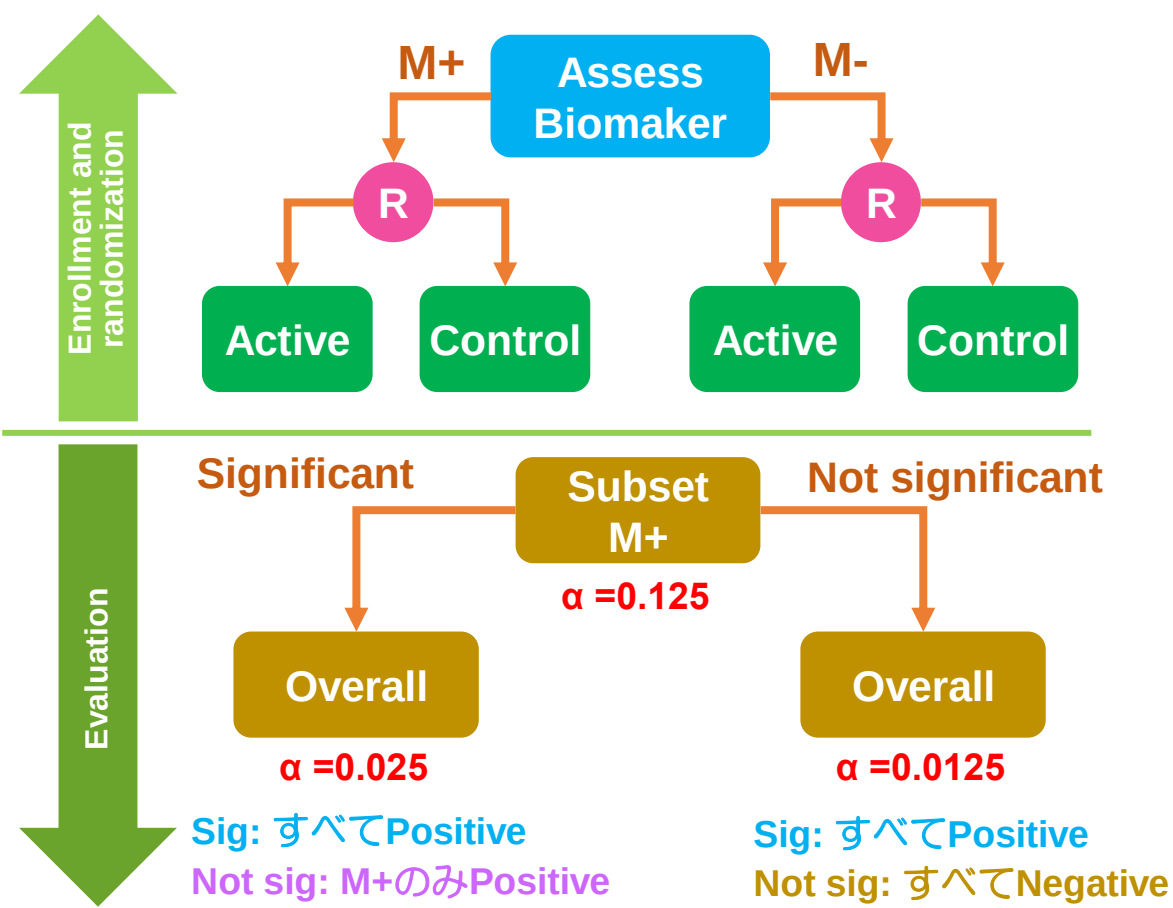
Design.3

Sequential Subgroup Specific Design



Design.4

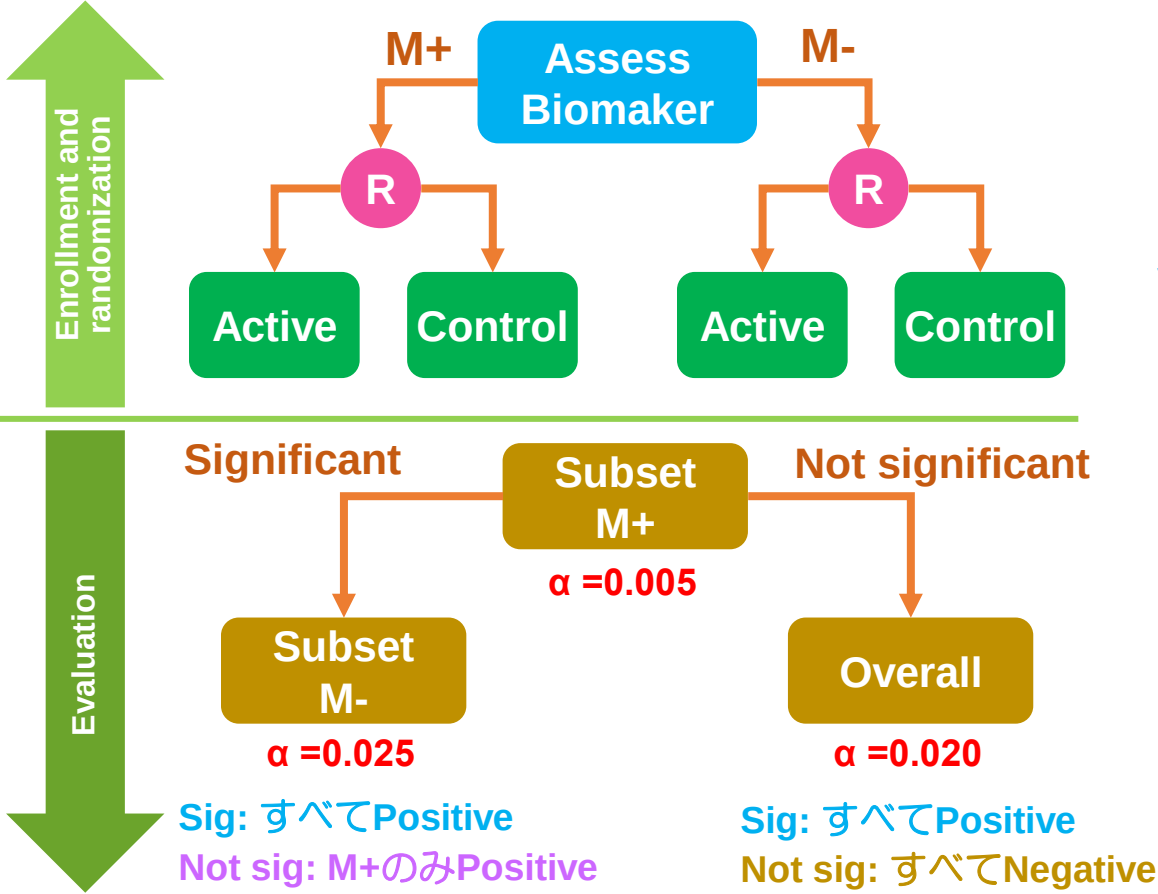
Biomarker-Positive and Overall Design



バイオマーカーを伴う試験

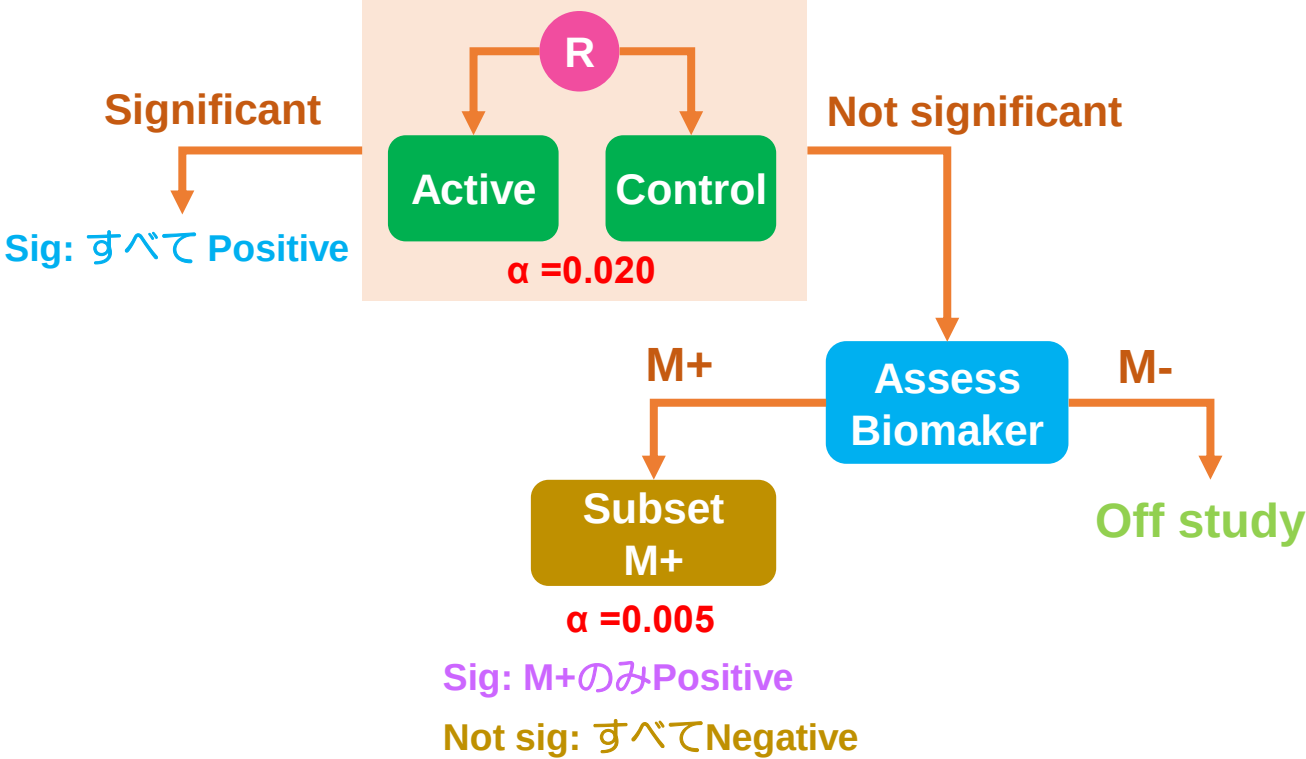
Design.5

Marker Sequential Test (MaST) Design



Design.6

Fall-back Design



バイオマーカーを伴う試験例

Design.1: Enrichment Design

ORIGINAL ARTICLE

Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation

Chapman et al. NEJM, 2011; 364:2507-2516.

BRAF V600E 変異を有する転移性悪性黒色腫患者に対する Vemurafenib(BRAF キナーゼ阻害薬)の無作為化比較
第III相試験

Primary endpoints

全生存期間(OS, 有意水準 $\alpha=0.045$), 無増悪生存期間(PFS, 有意水準 $\alpha=0.005$)

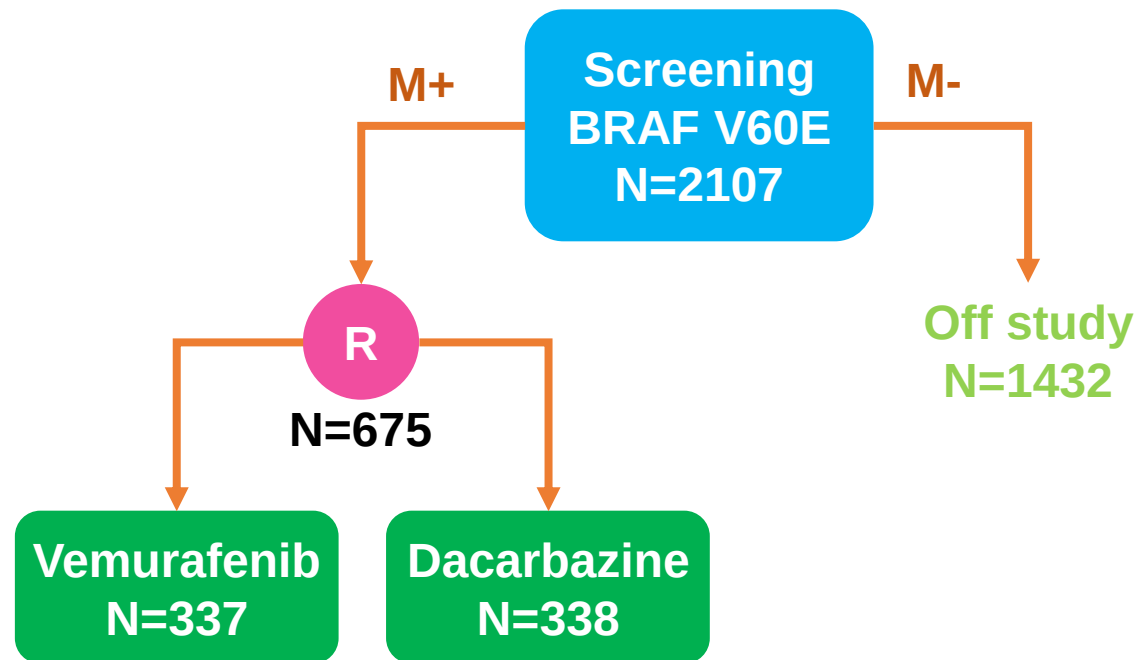
Sample size calculation

全生存期間: HR=0.65, 検出力 $1-\beta \geq 80\%$ → 196例のイベントが必要 → 680例(10%の不適合例を含む)

無増悪生存期間: HR=0.55, 検出力 $1-\beta \geq 90\%$ → 187例のイベントが必要

Result

OS HR=0.37 [0.26, 0.55] ($p<0.001^{*1}$), PFS HR=0.26 [0.20, 0.33] ($p<0.001^{*1}$)なので, BRAF V60E positive 症例に対するOS, PFSに対するVemurafenibの有効性が検証できた(*1: Stratified log-rank test).



バイオマーカーを伴う試験例

Design.3: Sequential Subgroup Specific Design

Randomized, Phase III Trial of Panitumumab With Infusional Fluorouracil, Leucovorin, and Oxaliplatin (FOLFOX4) Versus FOLFOX4 Alone As First-Line Treatment in Patients With Previously Untreated Metastatic Colorectal Cancer: The PRIME Study

Douillard et al. JCO, 2010; 28:4697-4705.

転移性大腸癌に対するPanitumumab+FOLFOX4 vs FOLFOX4
の無作為化比較第III相試験

Primary endpoints

無増悪生存期間(PFS, 有意水準 $\alpha=0.005$)

Sample size calculation

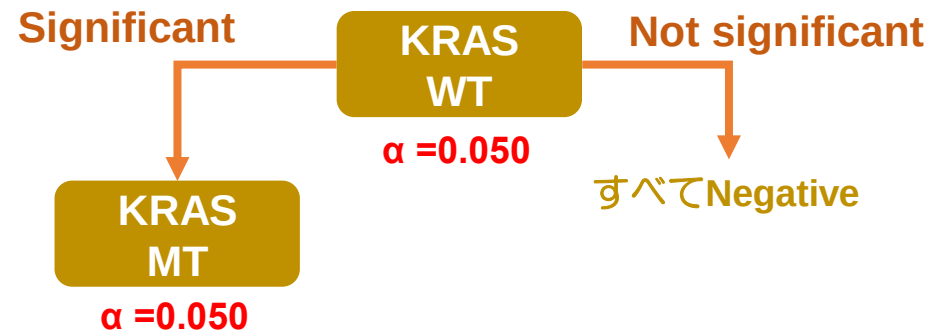
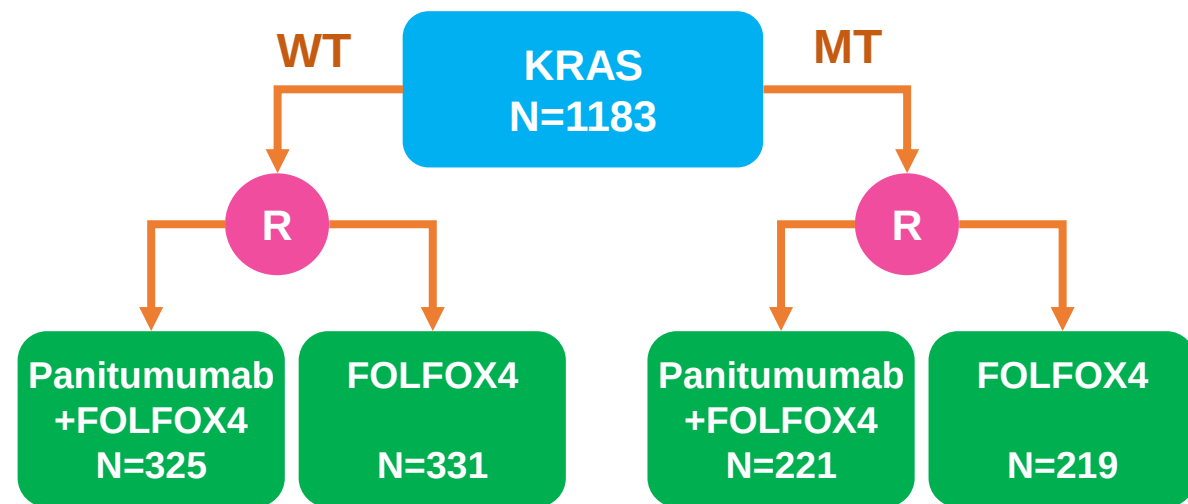
WT: HR=0.714, 検出力 $1-\beta \geq 90\%$

→ 380例のイベントが必要 (Total: 1150例)

MT: 症例設計なし

Result

WT OS HR=0.80 [0.66, 0.97] ($p<0.02^{*1}$), PFS MT=1.29 [1.04, 1.62] ($p<0.02^{*1}$)なので, KRAS WT/MTにおいて, 結果がクロスオーバーした(OSについても同様だがこちらは有意差なし)(*1: Stratified log-rank test).



Sig: すべてPositive

Not sig: M+のみPositive