# がん臨床試験における検定の多重性
## ～検定手順とグラフィカルアプローチ～

JCOGデータセンター統計部門
町田 龍之介

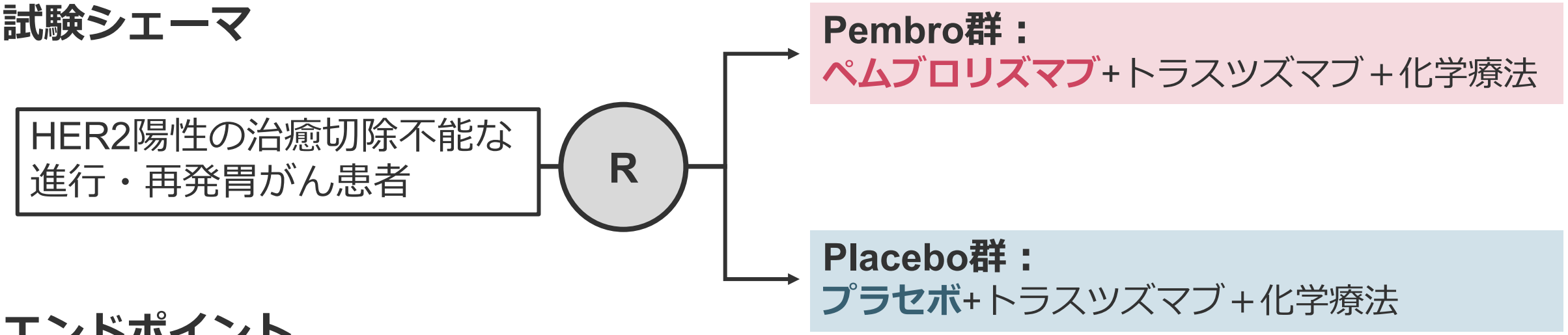
第28回JCOG臨床試験セミナー中級編

# 目次

- 実際の臨床試験における検定の多重性の例
- 検定の多重性の問題
- 基本的な多重比較法の概要
- グラフィカルアプローチ

# 実際の臨床試験の例（KEYNOTE-811）

## 試験シェーマ




## エンドポイント

Primary endpoints：無増悪生存期間 （PFS）と全生存期間（OS）
Secondary endpoint：奏効割合（ORR）など

## 主たる判断規準

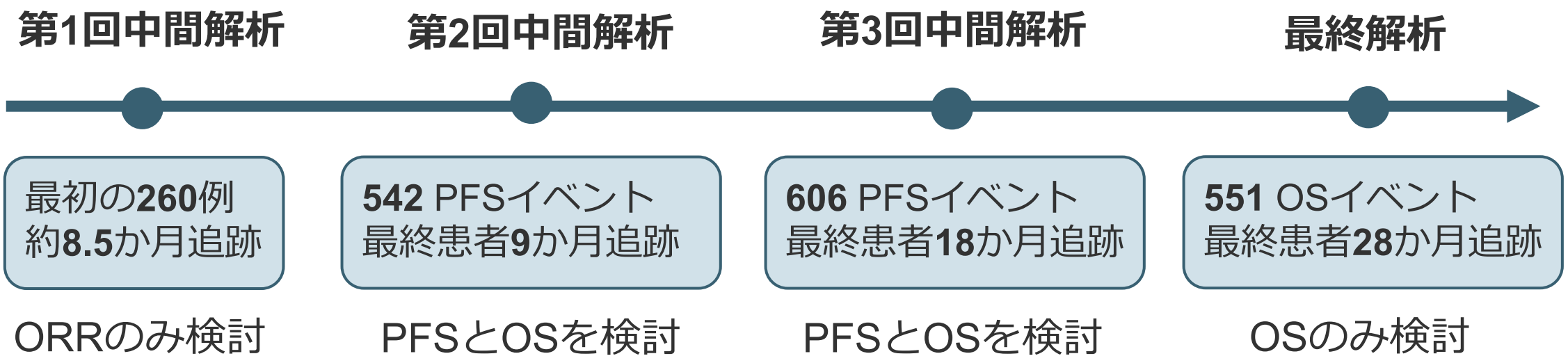
いずれかのprimary endpointでPembro群の優越性が検証されれば試験成功

Janjigian et al. (2023)

# 解析時期と検討するエンドポイント（計画）

- 登録数もしくはイベント数と追跡期間により解析時期を規定
- 解析時期により検討するエンドポイントが異なる

**予定サンプルサイズ：692例**




Janjigian et al. (2023)

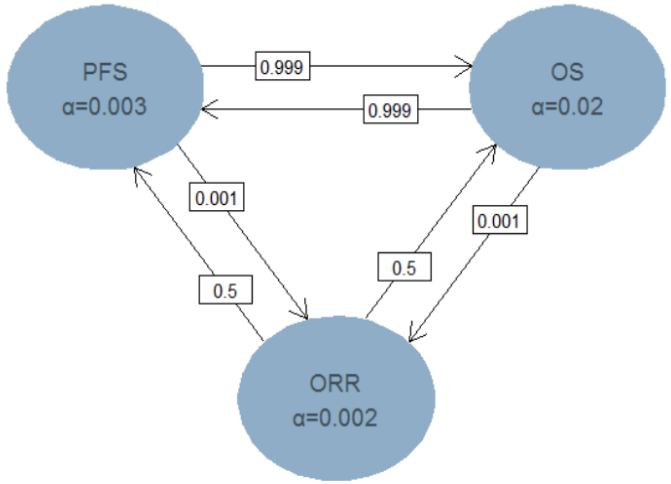
# 検定の多重性に関する統計記載（抜粋）

- 多重エンドポイントのα調整　グラフィカルアプローチを使用

An extension of the **graphical method** of Mauer and Bretz was used to control the overall type 1 error at a **one-sided α-level of 0·025**, with 0·002 originally allocated to objective response rate, 0·003 allocated to progression-free survival, and 0·02 to overall survival.

全体のα=0.025

Figure 4　Type I Error Reallocation Strategy




- 中間解析と最終解析間のα調整

有意水準はLan-DeMetsのα消費関数で更新

Statistical boundaries at each interim analysis were updated using a **Lan-DeMets O'Brien-Fleming spending fraction** on the basis of the information and α-level of the endpoint at each interim analysis.

Janjigian et al. (2023)

# 各解析時点のKEYNOTE-811の結果

## 第1回中間解析

ORR: Pembro群 74.4% vs. Placebo群 51.9%
**p=0.00006 < 有意水準α=0.002**

Janjigian et al. (2021)

## 第2回中間解析

PFSの優越性が示された
**p=0.0002 < 有意水準α=0.0013**

OSの優越性は示されず

Janjigian et al. (2023)

## 第3回中間解析

OSの優越性は示されず

## 最終解析

OSの優越性が示された
**p=0.0040 < 有意水準α=0.0201**

Janjigian et al. (2024)

# 本講義で想定するがん臨床試験

- 2群の検証的ランダム化比較試験（phase III）
- 3つのエンドポイント（ORR、PFS、OS）を設定
  - Key secondary endpoint（奏効割合）で有効性が認められた場合
    治療効果に関する付加的な情報をラベルに反映できる　FDA (2023)
- 最終解析前に中間解析を実施
- 試験全体のαエラーの確率を0.025に制御

**上記の仮想的な試験を想定して検定の多重性を理解する**

# 目次

- 実際の臨床試験における検定の多重性の例
- **検定の多重性の問題**
- 基本的な多重比較法の概要
- グラフィカルアプローチ

# 仮説検定の手順

## 群間に「差がある」ことを検証する方法

（差がある仮説を対立仮説と呼ぶ）

手順1：「差がない」という仮説（**帰無仮説**）を設定する

手順2：帰無仮説が正しい下で得られる結果の分布を算出する

手順3：実際に得られた結果と帰無仮説が正しい下での分布を比較して
どのくらいまれな結果なのかを確率（**p値**）で表す

手順4：p値が事前に設定した規準（**有意水準**）よりも低ければ、
最初に設定した帰無仮説が間違っていたと判断する

手順5：対立仮説「差がある」が正しいと結論づける

# 1つの帰無仮説に対する判断の誤り

- 仮説検定の結果は必ずしも正しくない
  - p値が有意水準以下（0.025や0.05）であれば有意差ありと判断
  - p値=0.004は仮に帰無仮説が正しくても0.4%は起こりうる結果

| | | 真実 | |
|---|---|---|---|
| | | 差なし | 差あり |
| 検定結果 | 有意差なし | 正しい | **第2種の過誤（$\beta$エラー）** |
| | 有意差あり | **第1種の過誤（$\alpha$エラー）** | 正しい（検出力） |

**$\alpha$エラーを起こす確率は最大でも有意水準以下になる**

# 複数の帰無仮説に対する判断の誤り

- 1つ以上の正しい帰無仮説（真に差がない仮説）を棄却してしまう確率
  - この確率は**Family-wise error rate（FWER）**と呼ばれる

| 仮説の数 | FWER* |
|---|---|
| 1 | 2.5% |
| 2 | 4.9% |
| 3 | 7.3% |
| 5 | 11.9% |
| 10 | 22.4% |
| 20 | 39.7% |
| 100 | 92.0% |

* 有意水準0.025で各結果が独立の場合

FWER = $1-(1-0.025)^m$ , m: 仮説の数

いつくかのαエラーの確率の指標が提案されているが、検証的試験では1つの仮説に対しても誤った判断をすることは望ましくないためFWERを制御する検定手法が用いられる

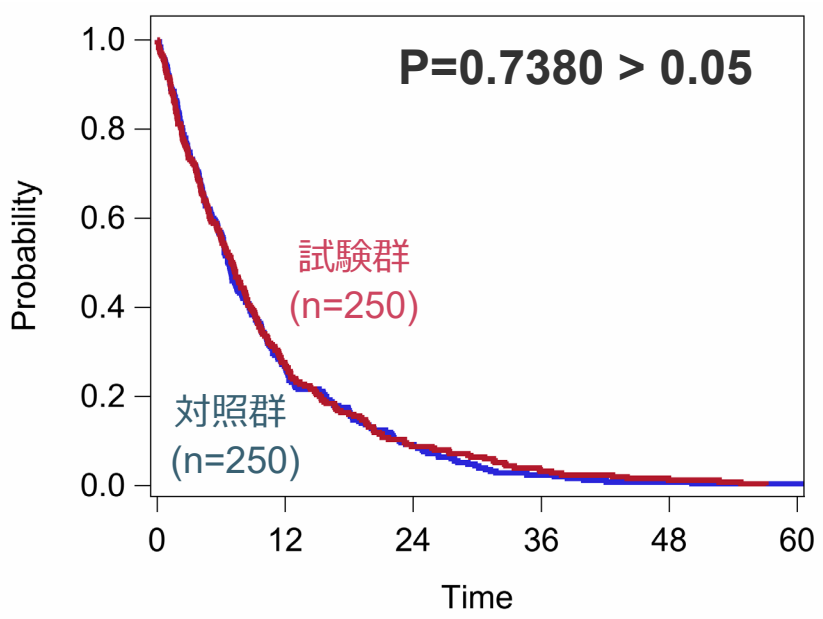
真に差がない10個の仮説を検定すれば約20%の確率で誤った判断が含まれる

# 検定の多重性とは

■ “都合の良い”結果を選択する際に生じるFWERの増加

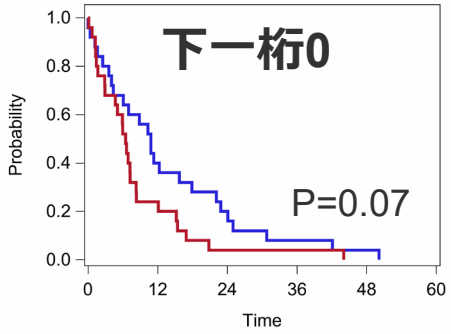
## 全体の群間比較の結果

「差がない」が真実の下での仮想的な結果

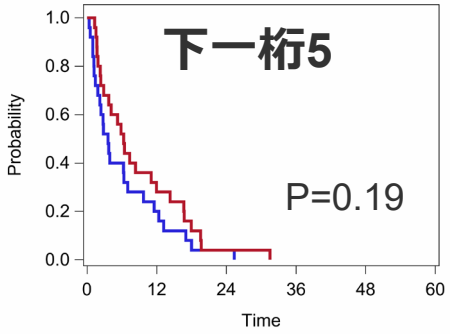


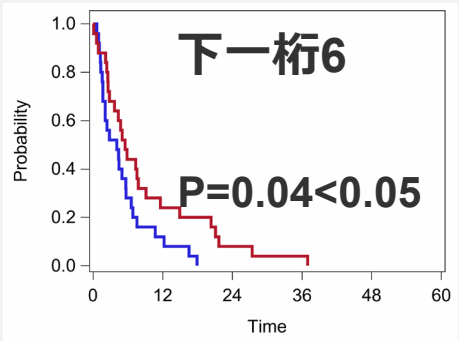

## サブグループ解析の結果

登録番号の下一桁で分けた10個のサブグループ



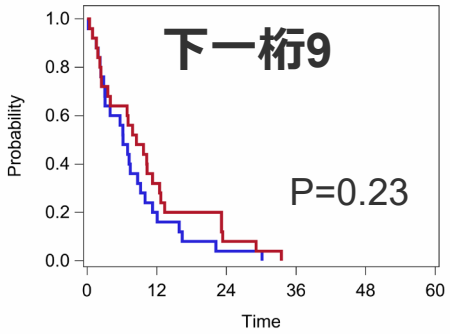

（中略）




偶然差が見られる

（中略）




参考：ICRweb 柴田先生「検証的試験における 多重性の調整」

# がん臨床試験における検定の多重性

- **多重エンドポイント（multiple endpoints）**
  - 複数のprimary endpointやsecondary endpointを検証

- **多群比較（multi-arm comparison）**
  - 1つの対照治療群と2つ以上の試験治療群を比較

- **群逐次試験（group sequential trial）**
  - 異なる時期に複数回解析（中間解析）を実施

# 目次

# 検定の多重性の調整法

**がん臨床試験で良く用いられる3つの方法**
**（p値に基づく検定手順）**

- αの再利用・再分配（alpha recycling or re-allocation）
- αの分割・配分（alpha splitting or allocation）
- 両方を合わせたαの調整の方法

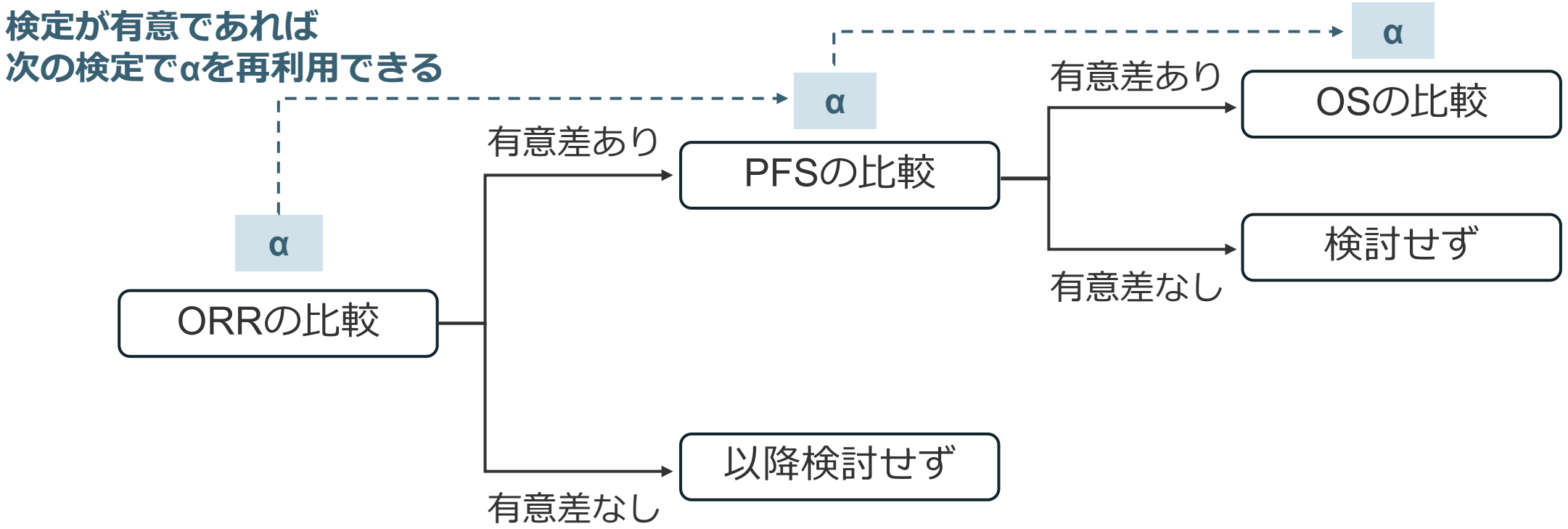
※ 統一的にORR、PFS、OSを例に説明しているが
本状況において各方法を必ずしも推奨している訳ではない

# 固定順序手順（検定順序を固定）

Maurer et al. (1995)

■ 事前に規定した順番で各検定を有意水準αで検定する

**例）ORR→PFS→OSの順に検定**

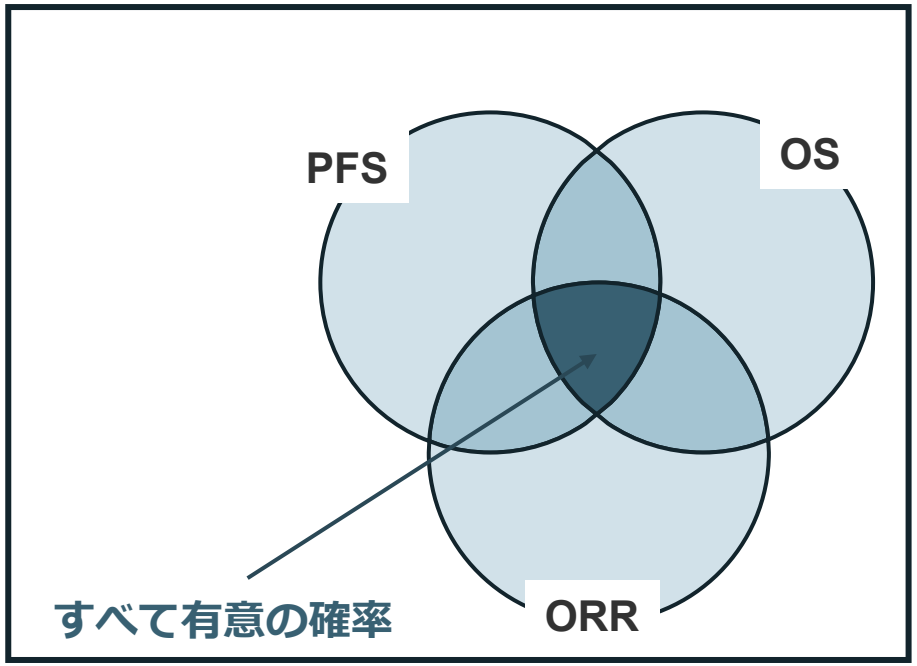



**前の検定が有意差なしの場合、次の検定を行えない**

# 本当に有意水準α以下に制御される？

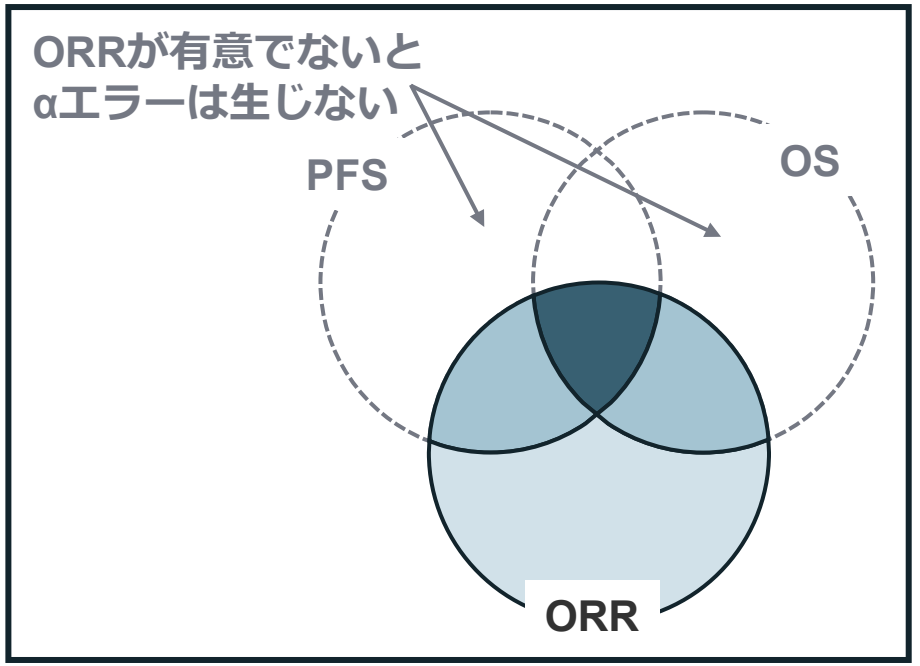
■ すべての帰無仮説が正しい（**差がない**）場合

**検定の順番に関係なく**各検定を
αで検定したときの**αエラーの確率**




αエラーの確率は**最大で3α**
（それぞれの円の確率はα）

**ORR→PFS→OSの順に**各検定を
αで検定した場合の**αエラーの確率**

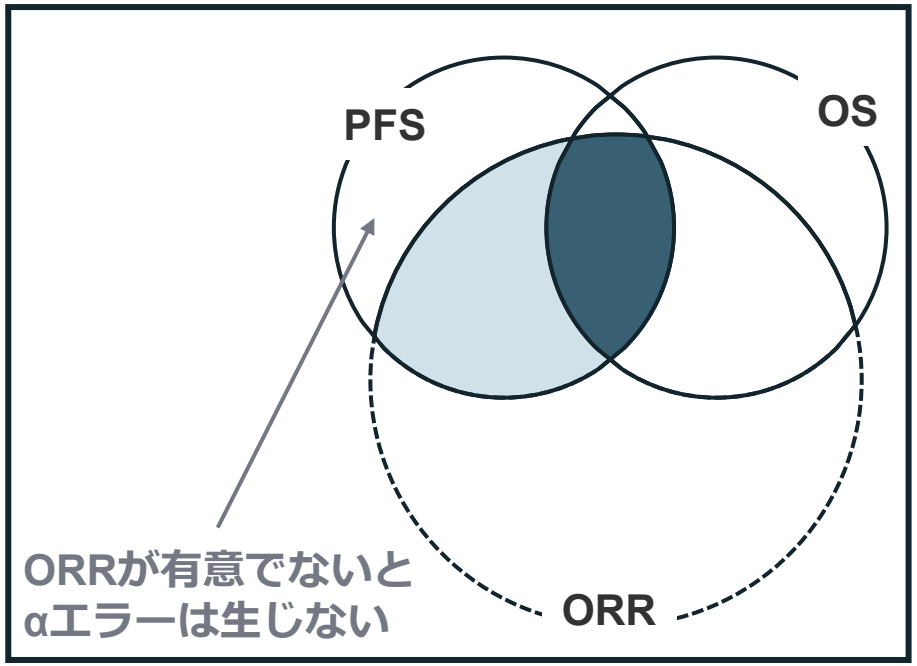



αエラーの確率は**最大でα**

# 帰無仮説の正誤が異なる場合は？

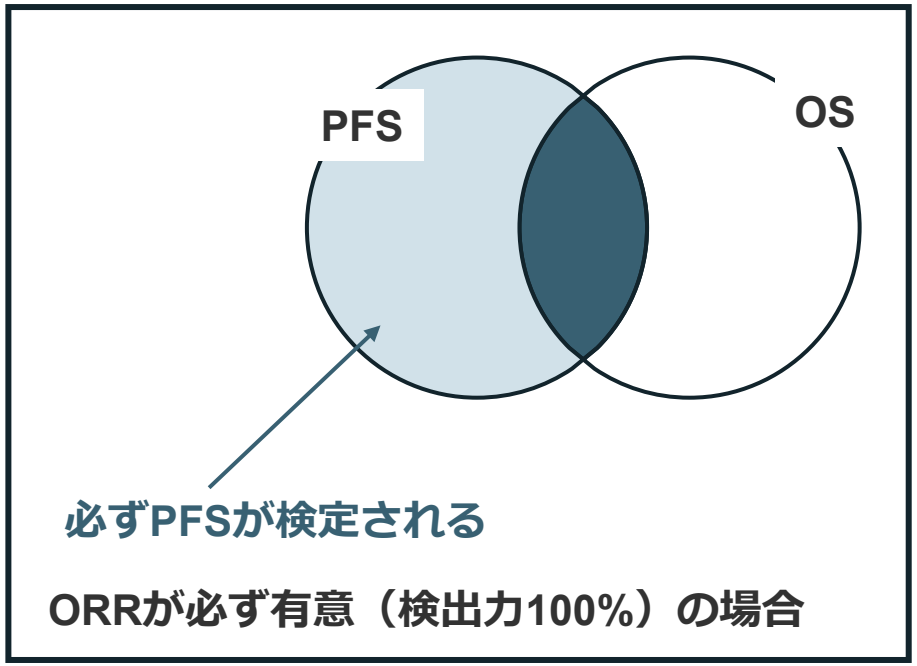
■ ORRの帰無仮説は誤り（**差あり**）でPFSとOSは正しい（**差がない**）場合

**ORR→PFS→OSの順に**各検定を
αで検定した場合の**αエラーの確率**




αエラーの確率は**最大でもα**

**ORR→PFS→OSの順に**各検定を
αで検定した場合の**αエラーの確率**

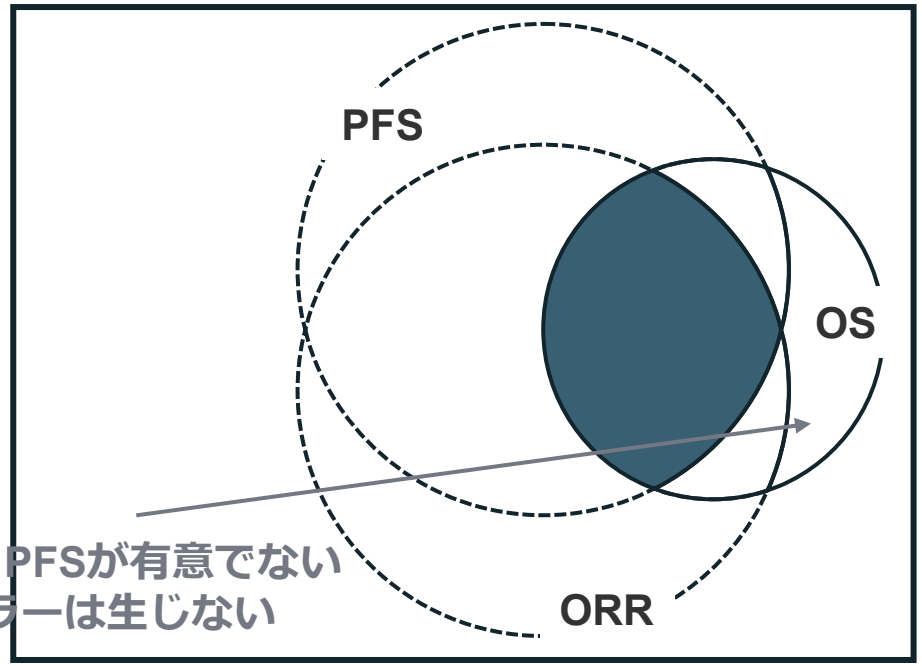



極端な状況を仮定しても
αエラーの確率は**最大でα**

# 帰無仮説の正誤が異なる場合は？

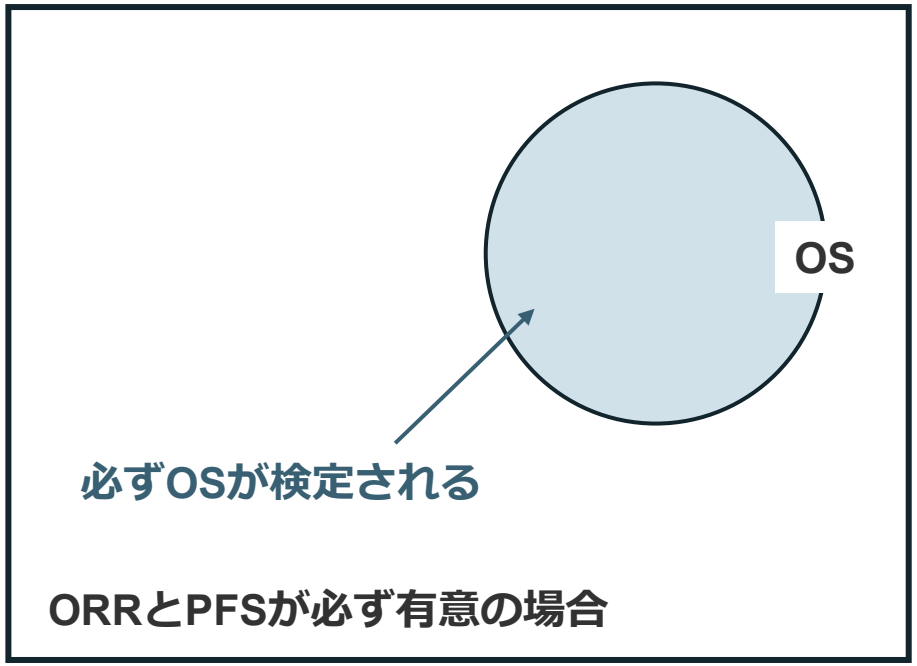
■ ORRとPFSの帰無仮説は誤り（**差あり**）でOSは正しい（**差がない**）場合

**ORR→PFS→OSの順に**各検定を
αで検定した場合の**αエラーの確率**




αエラーの確率は**最大でもα**

**ORR→PFS→OSの順に**各検定を
αで検定した場合の**αエラーの確率**




極端な状況を仮定しても
αエラーの確率は**最大でα**

# 固定順序手順のαエラー

αエラー：正しい帰無仮説を誤っていると判断する確率

- 8つの帰無仮説の正誤のパターン（$2^3$=8通り）

| | ORR | PFS | OS |
|---|---|---|---|
| 1 | **正（差なし）** | 正（差なし） | 正（差なし） |
| 2 | | | 誤（差あり） |
| 3 | | 誤（差あり） | 正（差なし） |
| 4 | | | 誤（差あり） |
| 5 | 誤（差あり） | **正（差なし）** | 正（差なし） |
| 6 | | | 誤（差あり） |
| 7 | | 誤（差あり） | **正（差なし）** |
| 8 | | | 誤（差あり） ← |

正しい帰無仮説はないためαエラーは生じない

- ORR→PFS→OSの順で**最初に検定される正しい帰無仮説（差なし）**で使用した有意水準以下にαエラーの確率は制御される

# 2つ意味でのFWERの制御

**FWERを強い意味で制御（in the strong sense）**

**すべての帰無仮説の正誤のパターンでFWERがα水準以下**

**FWERを弱い意味で制御（in the week sense）**

**すべての帰無仮説が正しいときにFWERがα水準以下**

**真の状況は不明でありFWERを強い意味で制御する方法が要求される**

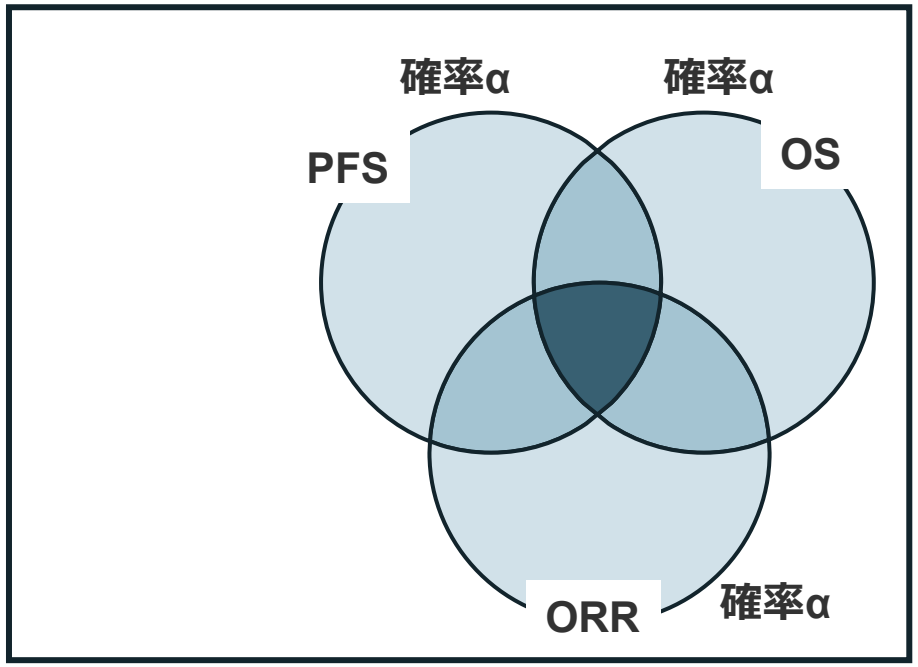
（以降はFWERを強い意味で制御する方法を紹介する）

# Bonferroni手順（αを均等に分割）

- 全体の有意水準αを検定回数で割った値を各検定に使用

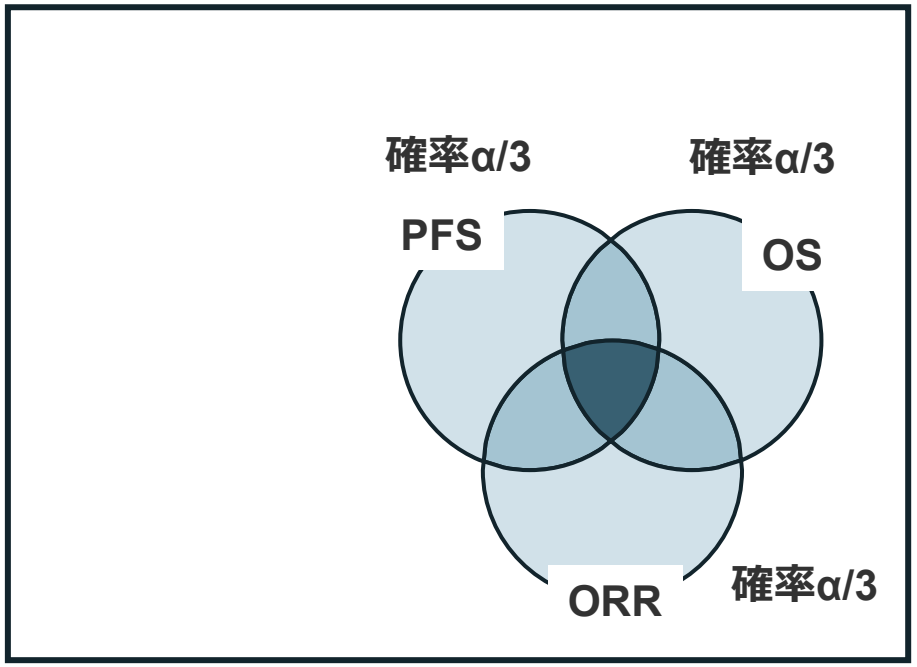
**例）ORR、PFS、OSの3つの検定にα/3ずつ分割**

αで検定したときの**αエラーの確率**




αエラーの確率は**最大で3α**

**α/3**で検定したときの**αエラーの確率**



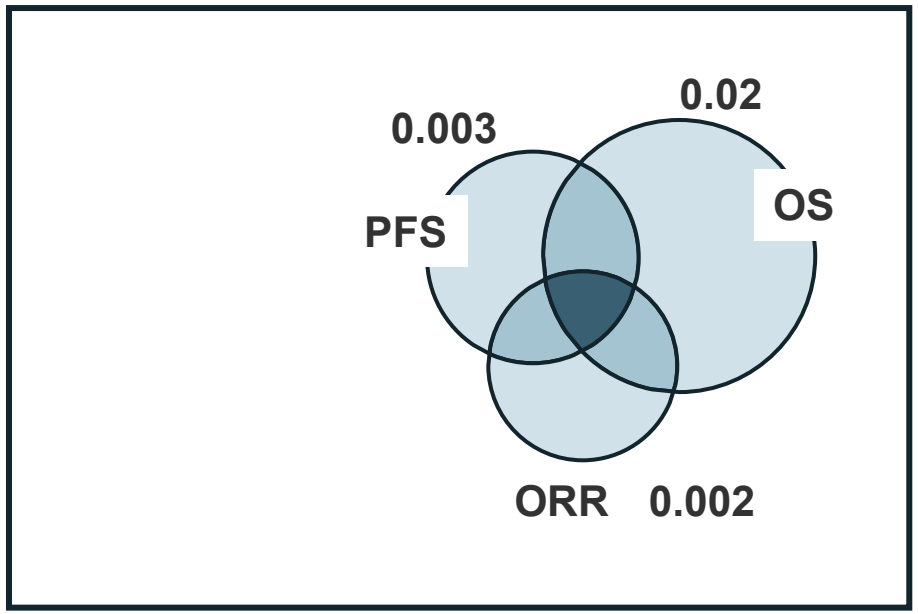

αエラーの確率は**最大でα**
（完全に円の重なりがない場合）

# 重み付きBonferroni手順

- 各検定の有意水準の合計がαとなるように分割

**例）ORRに0.002、 PFSに0.003、 OSに0.02を分割（KEYNOTE-811）**

**すべての帰無仮説が正しい（差がない）場合のαエラーの確率**




**（重み付き）Bonferroni手順のまとめ**

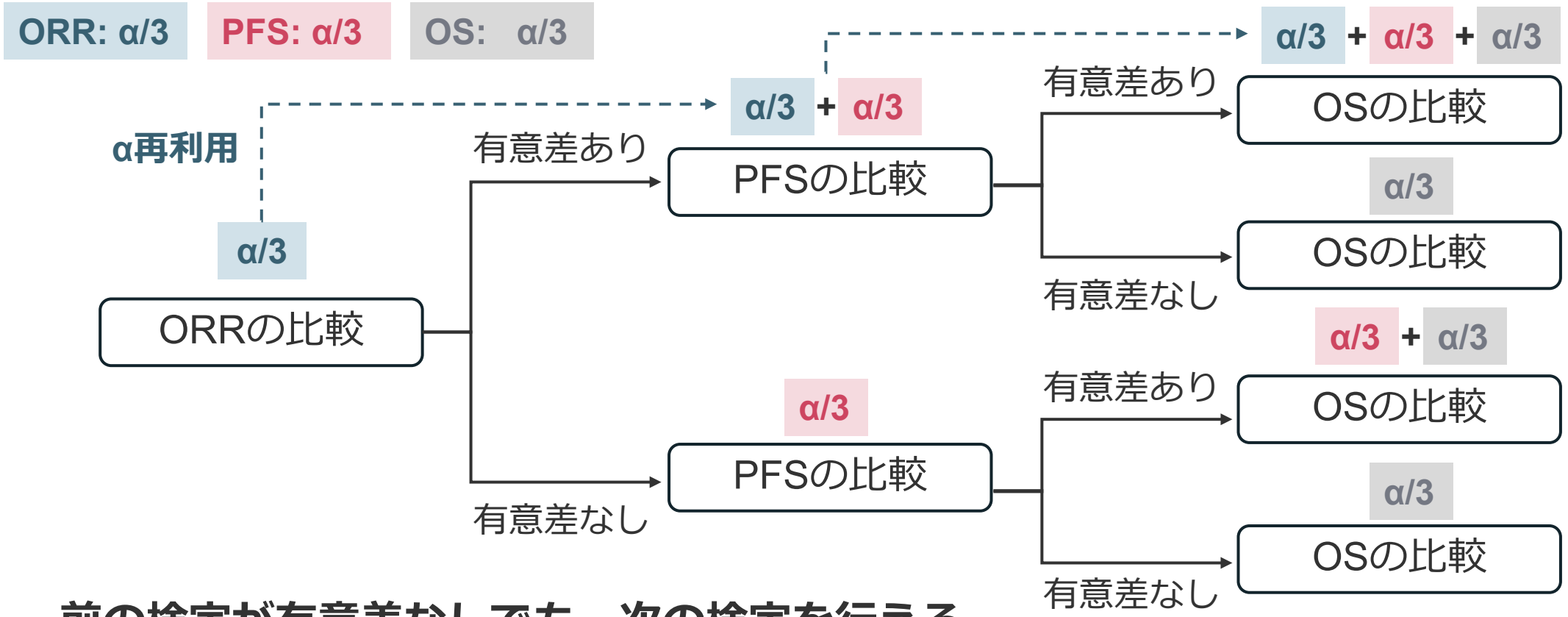
- 固定順序手順とは異なり **すべての帰無仮説を検討**できる
- 他の検定結果に依存しないため **臨床的な解釈がシンプル**
- 正しい帰無仮説が同時に棄却される状況を考慮しておらず**保守的である**（検定で有意になりづらい）

# Fallback手順（α分割あり・順序固定あり）

Wiens (2003), Wiens and Dmitrienko (2005)

- 最初に各検定にαを分割し、検定の順番を決定

**例）各検定にα/3ずつ分割し、ORR→PFS→OSの順に検定**



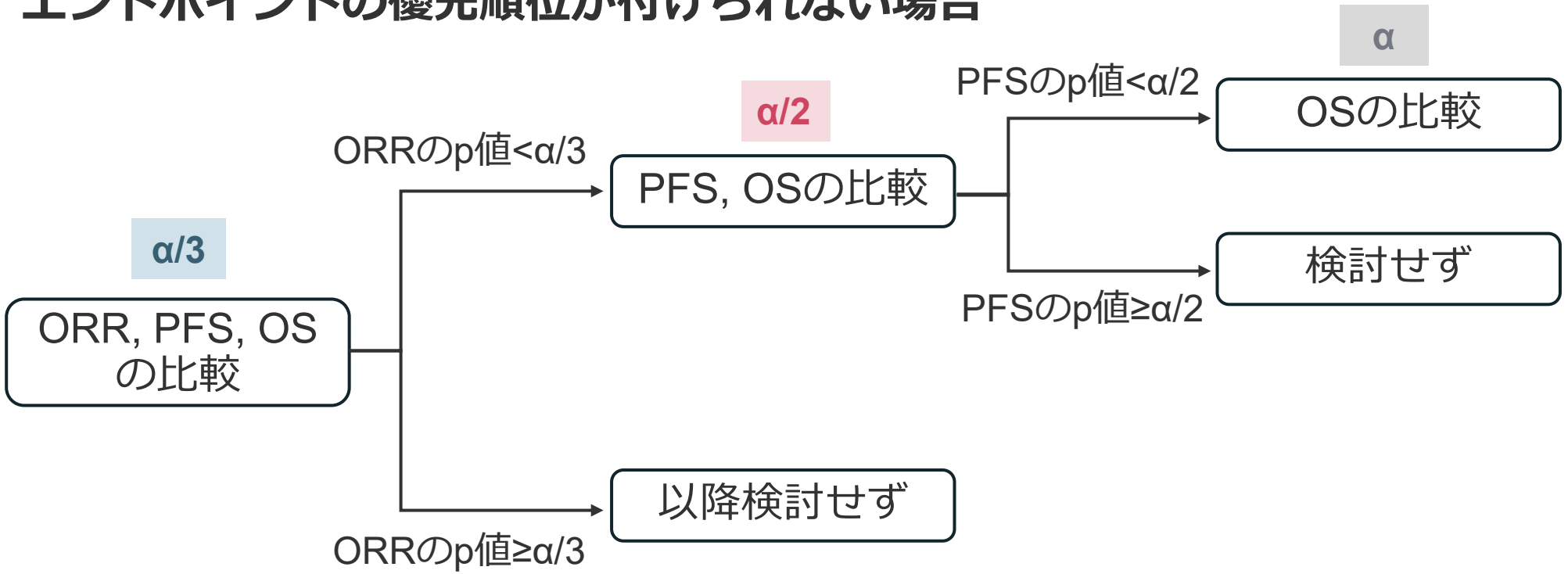

**前の検定が有意差なしでも、次の検定を行える**

# Holm手順（α分割なし・順番固定なし）

Holm (1979)

- 検定の順番はデータ（実際のp値）に応じて決定

**例）ORR→PFS→OSの順にp値が小さいとする**
**エンドポイントの優先順位が付けられない場合**




**前の検定が有意差なしの場合、次の検定を行えない**

# 多重性の調整の検定手順まとめ

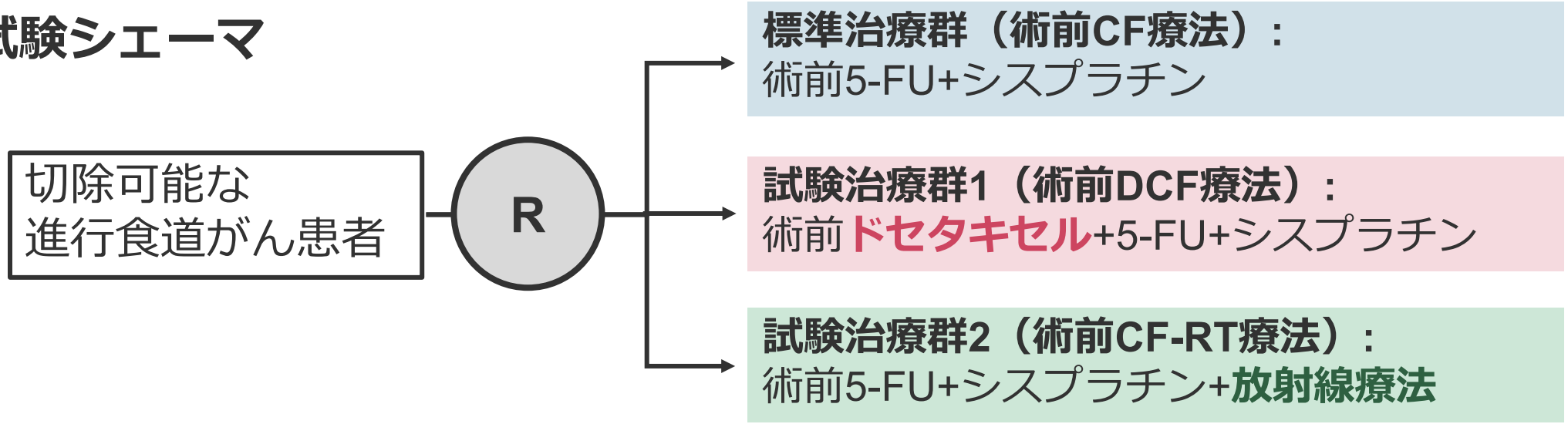
| 手順 | 検定の順番 | α分割 | α再利用 |
|---|---|---|---|
| 固定順序手順 | 事前に固定 | なし | あり |
| Bonferroni手順 | 1回の検定のみ<br>(全検定を検討可能) | あり | なし |
| Fallback手順 | 事前に固定<br>(全検定を検討可能) | あり | あり |
| Holm手順 | データに応じて決定 | (なし) | あり |

**臨床的な作用機序や試験の目的に適合した検定手順を選択**

# Bonferroni手順の実例（JCOG1109）

Kato et al. (2024)

## 試験シェーマ




## 目的

術前CF療法に対する術前DCF療法と術前CF-RT療法のOSにおける優越性

## 多群比較による多重性の調整（抜粋）

全体のαを5%に抑えるためBonferroni法を使用

We adjusted for multiplicity due to two pairwise comparisons **with the Bonferroni method** to maintain **the study-wise one-sided α level of 5%** (ie, a one-sided **α level of 2·5% was assigned** to each paired comparison [NeoCF *vs* NeoCF+D or NeoCF *vs* NeoCF+RT]).

各対比較に2.5%ずつαを分割

# 3群比較の多重性をHolm手順で調整？

**Bonferroni手順で各比較に0.025ずつ分割（全体のα=0.05）**

| | | | | |
|---|---|---|---|---|
| 比較1 | 術前CF療法 | vs. | 術前DCF療法 | p=0.006 < 0.025*（有意差あり） |
| 比較2 | 術前CF療法 | vs. | 術前CF-RT療法 | p=0.12 > 0.025*（有意差なし） |

*簡単のため時点間の多重性は考慮せず

**仮にHolm手順で多重性を調整していたら…**

| | | | | |
|---|---|---|---|---|
| 比較1 | 術前CF療法 | vs. | 術前DCF療法 | p=0.006 < 0.025（有意差あり） |
| 比較2 | 術前CF療法 | vs. | 術前CF-RT療法 | p=0.12 > 0.05（有意差なし） |

**JCOG1109の場合は結果は変わらず**

- 仮に比較2のp値が0.025以上0.05未満の場合、Bonferroni手順と判断が異なる
- **片方の検定結果がもう一方の検定結果に左右されることは臨床的に不自然**

Kato et al. (2024)

# 目次

- 実際の臨床試験における検定の多重性の例
- 検定の多重性の問題
- 基本的な多重比較法の概要
- **グラフィカルアプローチ**

# グラフィカルアプローチの基本

Bretz et al. (2009)

## 検定手順を視覚的に表現する方法

- **手順1：各帰無仮説（ノード）を円で表現**

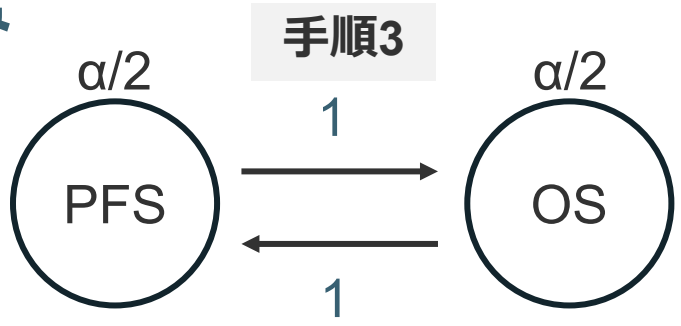
  例）PFSとOSの2つの帰無仮説を設定


- **手順2：合計でαになるように各検定に分割**

  例）PFSとOSの帰無仮説にα/2ずつ

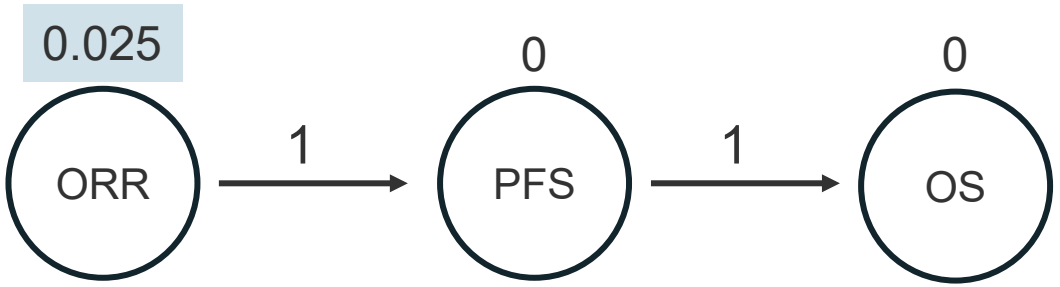

- **手順3：ある帰無仮説が棄却された時の再利用の重み**
  - 各ノードからの重みの合計は最大で1（100%）

  例）Holm手順のように片方が有意の場合
  もう一方に100%再利用（Holm手順）



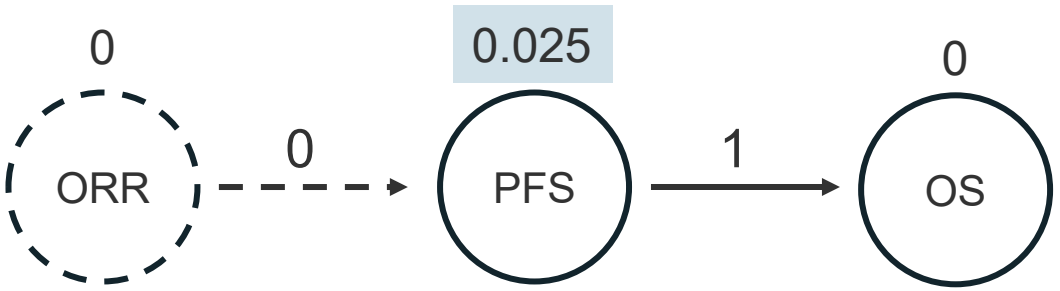
# 固定順序手順の視覚的表現

## 初期グラフ

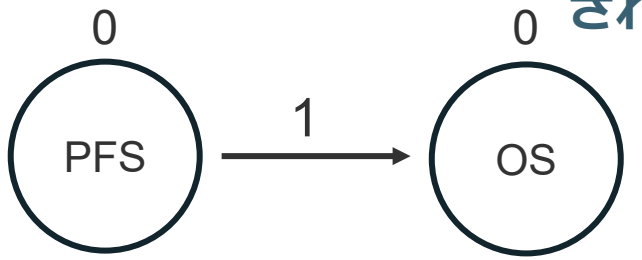



ORRが有意でないとPFSとOSは
検定されないため初期グラフでのαは0

## ORRが有意の場合

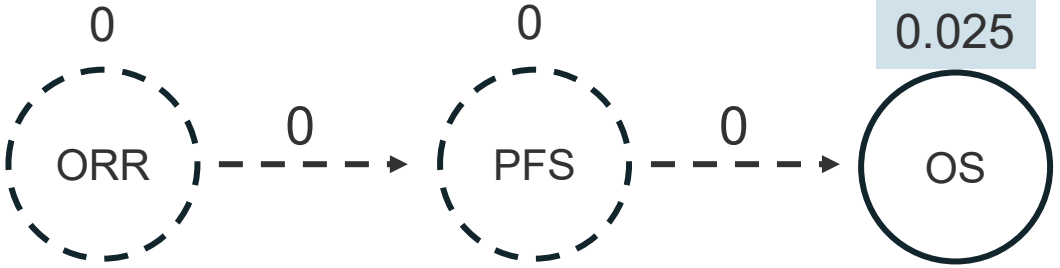



## ORRが有意でない場合

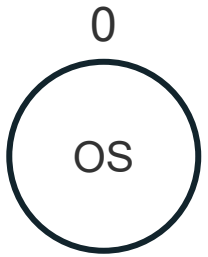



**PFSとOSにαが再利用
されず検定できない**

## ORRとPFSが共に有意の場合




## ORRが有意、PFSが有意でない場合



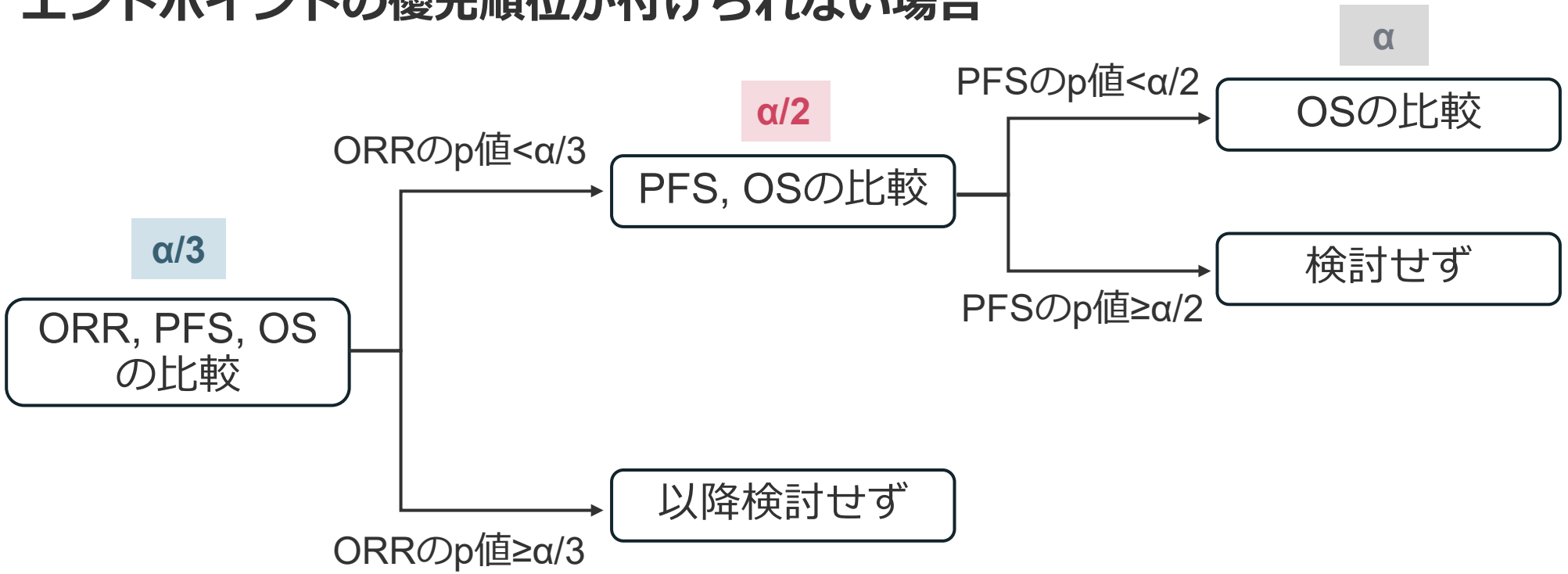

**PFSのαがOSに再利用
されず検定できない**

# 再掲：Holm手順（α分割なし・順番固定なし）

Holm (1979)

■ 検定の順番はデータ（実際のp値）に応じて決定

**例）ORR→PFS→OSの順にp値が小さいとする**
**エンドポイントの優先順位が付けられない場合**

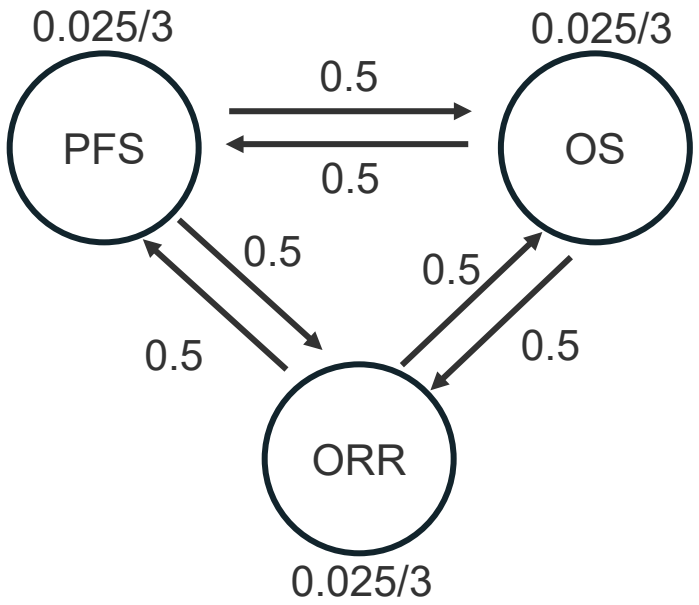



**前の検定が有意差なしの場合、次の検定を行えない**

# Holm手順の視覚的表現

## 例）ORR→PFS→OSの順にp値が小さいとする

### 初期グラフ



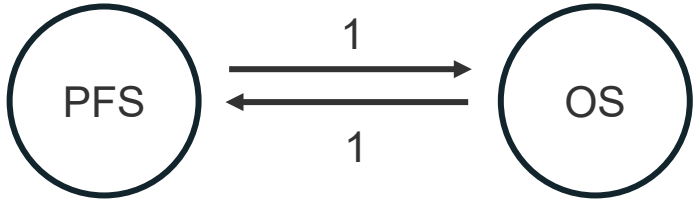

・最初にORR、PFS、OSそれぞれ0.025/3ずつ分割
・ある仮説が棄却されたら残りの仮説に等分にα分配

**前のスライドの有意水準と一致するのか？**

### ORRが有意の場合

**0.025/2**
(0.025/3+0.025/3×0.5)

**0.025/2**
(0.025/3+0.025/3×0.5)




ORRに配分予定だった重みを
他の検定の重みの比率で再計算

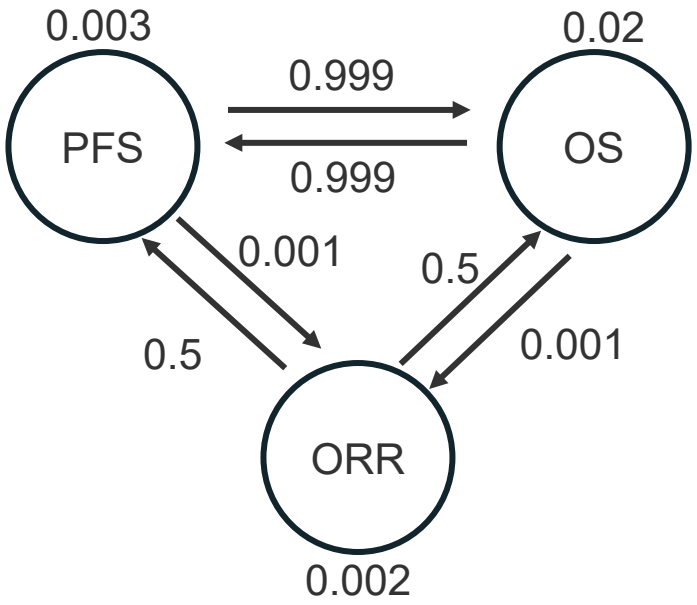
### ORRとPFSが共に有意の場合

**0.025**
(0.025/2+0.025/2)

# グラフィカルアプローチの実例（KEYNOTE-811）

Janjigian et al. (2023)

## 初期グラフ

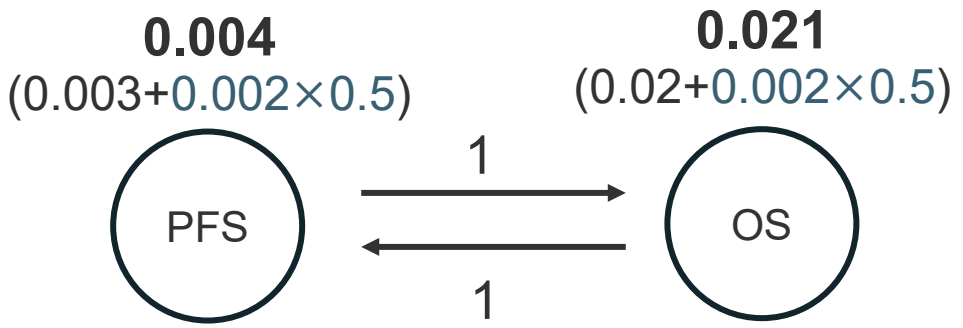



※統計記載を基に独自に作成

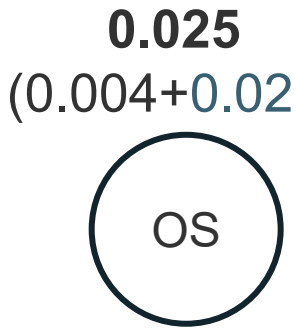
第1回中間解析でORRをα=0.002で検定して有意でない場合にはPFSとOSで有意となった後に再度ORRの検定を行う予定であった

## 第2回中間解析時のグラフ

**第1回中間解析でORRが有意**
**ORRのαがPFSとOSに均等に再分配**




**さらに第2回中間解析でPFSが有意**
**PFSのαがOSに100%再分配**



# クイズ：ORRが有意になった場合

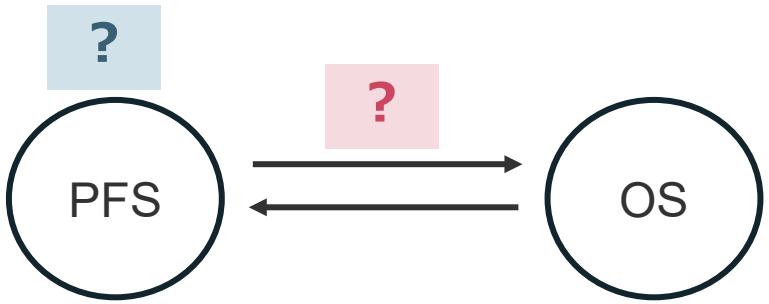
**問題1**：更新後のグラフのPFSのαは？

① 0.025/3+0.025/3　② 0.025/2　③ 0.025

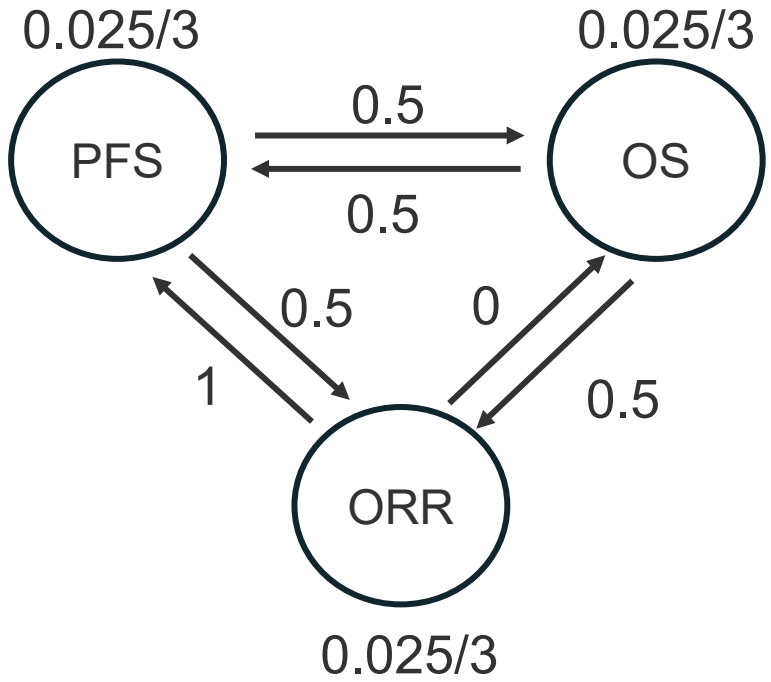
**問題2**：更新後のグラフのPFSのOSへの重みは？

① 0.25　② 0.5　③ 1

**更新後のグラフ**

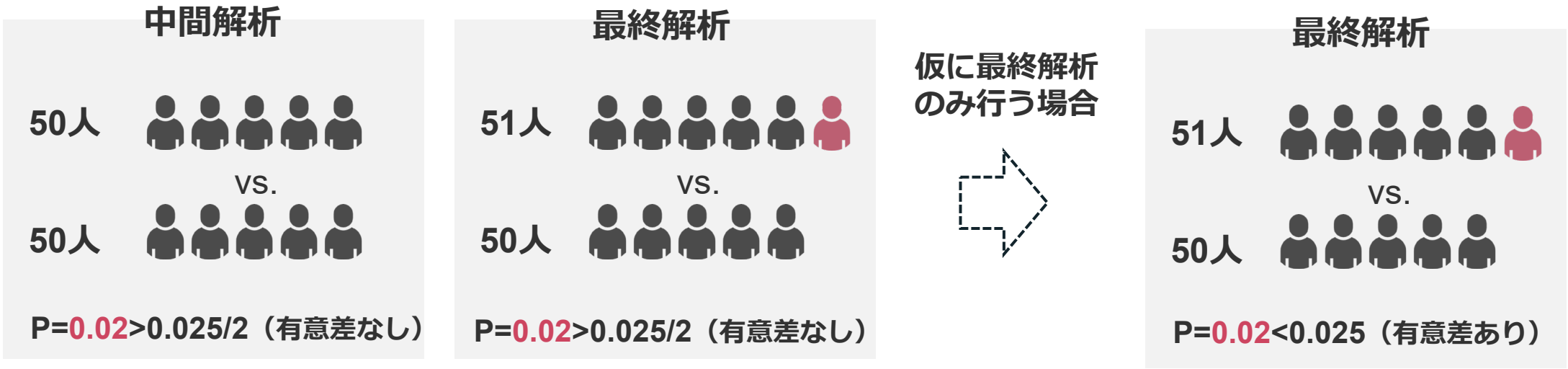



**初期グラフ**



# 群逐次試験における検定の多重性

- 中間解析と最終解析で複数回検定を行い、いずれかの時点で帰無仮説が棄却されれば試験治療が有効と判断するため、検定の多重性の問題が生じる
- 中間解析のデータは最終解析のデータにも含まれる（相関がある）




- Bonferroniの方法でもαを制御できるが有意になりづらく保守的である

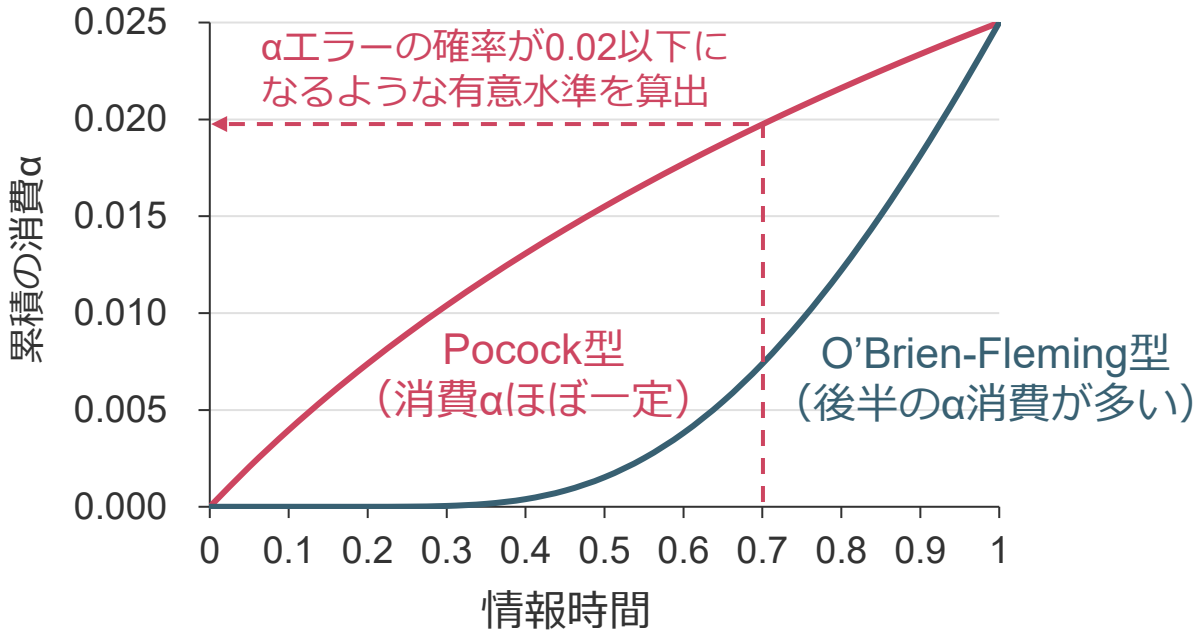
**中間解析と最終解析のデータが似ていることを考慮した調整法は？**

参考：ICRweb 水澤先生「中間解析の基礎と実際の運用」

# Lan and DeMetsのα消費関数による調整

Lan and DeMets (1983)

- 中間解析時の最終解析時の結果の相関を情報時間で表して、各解析時点の有意水準を計算する方法
  - 情報時間 = 各時点の**登録数**/予定の登録数（連続、二値アウトカム）
  - 情報時間 =各時点の**観察イベント数**/必要イベント数（事象時間アウトカム）




**各時点の有意水準**

| 情報時間 | Pocock型 | O-F型 |
|---|---|---|
| 0.7<br>（第1回中間解析） | 0.020 | 0.007 |
| 0.9<br>（第2回中間解析） | 0.011 | 0.016 |
| 1.0<br>（最終解析） | 0.009 | 0.019 |

どちらも合計が0.025以上となり
Bonferroniの方法よりも効率的

# 多重エンドポイントを設定した群逐次試験

**多重エンドポイントと時点間の多重性を同時に調整する必要がある**

- 多重エンドポイントの多重性の調整
  - 固定順序手順（**検定に順番を付けたい場合**）
  - Bonferroni手順（**互いの検定結果に影響されないようにしたい場合**）
  - Fallback手順（**検定に順番を付けるがすべて検討したい場合**）
  - Holm手順（**検定の順番を付けられない場合**）

- 時点間の多重性の調整法
  - O’Brien-Fleming型のα消費関数（**JCOGのデフォルト**）
  - Pocock型のα消費関数

# 群逐次試験におけるPFSとOSの多重性

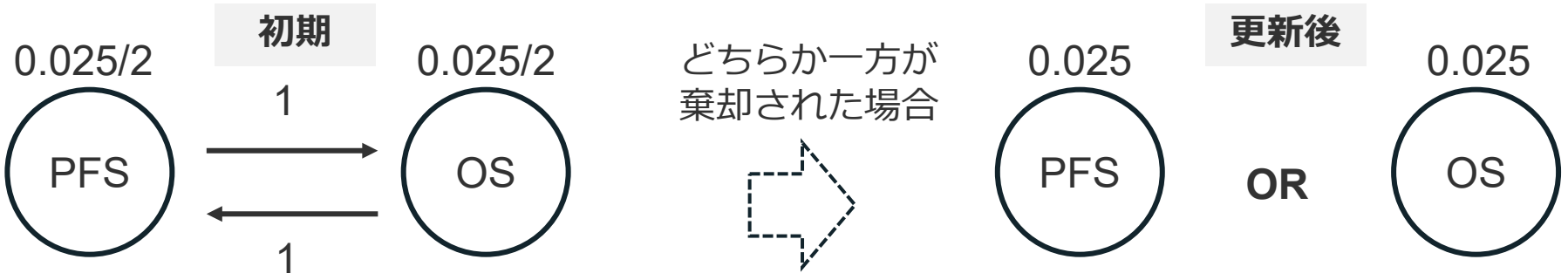
Maurer and Bretz (2013)

**仮想例**：第1、2回中間解析は**PFSの情報時間が0.9と1**の時点（OSのイベント数は問わない）で実施し、最終解析は**OSの情報時間が1**の時に実施する臨床試験

| | 各解析時点の情報時間 | | |
|---|---|---|---|
| | 第1回中間解析 | 第2回中間解析 | 最終解析 |
| PFS | **0.90** | **1.00** | ― |
| OS | 0.70 | 0.90 | **1.00** |

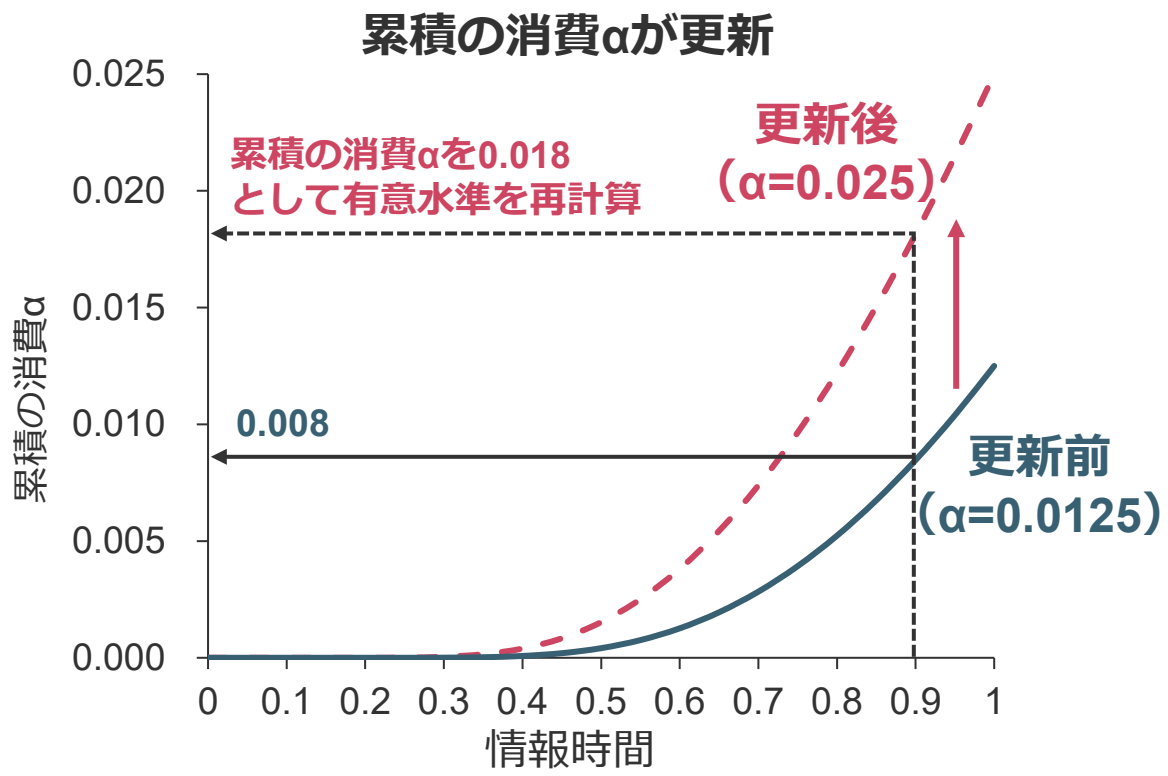
情報時間0.9=必要イベント数の90%のイベント数が観察

- 全体のαを0.025に制御するため各エンドポイントに0.0125ずつ分割（**Holm手順**）




- 両方のエンドポイントとも時点間の多重性を**O'Brien-Fleming型**のα消費関数で調整

# 更新後のα消費関数と有意水準




例）第2回中間解析でPFSが有意の場合
更新前：OSの**p値<0.008**で有意
更新後：OSの**p値<0.016**で有意

## PFSで使用する有意水準

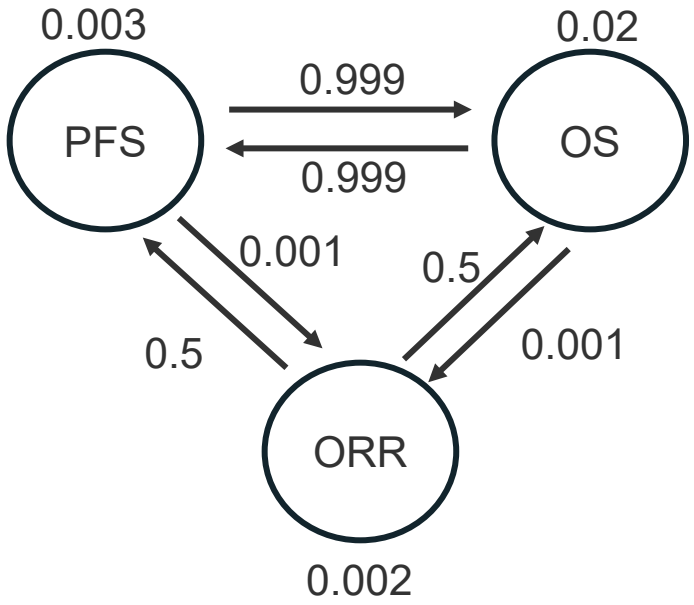
| 情報時間 | α=0.0125 | α=0.025 |
|---|---|---|
| 0.9<br>（第1回中間解析） | 0.008 | 0.018 |
| 1.0<br>（第2回中間解析） | 0.010 | 0.020 |

## OSで使用する有意水準

| 情報時間 | α=0.0125 | α=0.025 |
|---|---|---|
| 0.7<br>（第1回中間解析） | 0.003 | 0.007 |
| 0.9<br>（第2回中間解析） | 0.008 | 0.016 |
| 1.0<br>（最終解析） | 0.010 | 0.019 |

# OSの各時点の有意水準（KEYNOTE-811）

- OSの検定前にはORRとPFSが有意か否かの4パターンが考えられる
- 計画段階では第2、3回中間解析、最終解析で**予想される情報時間**を基に全4パターンの各解析時点における有意水準が検討されていた




| 解析時期/<br>情報時間 | ORR 差なし<br>PFS 差なし<br>**α=0.02** | ORR 差あり<br>PFS 差なし<br>**α=0.021** | ORR 差なし<br>PFS 差あり<br>**α=0.023** | ORR 差あり<br>PFS 差あり<br>**α=0.025** |
|---|---|---|---|---|
| 第2回中間解析<br>73%（401/551） | 0.0063 | 0.0068 | 0.0076 | 0.0085 |
| 第3回中間解析<br>89%（401/551） | 0.0115 | 0.0122 | 0.0134 | 0.0147 |
| 最終解析<br>100%（551/551） | 0.0157 | 0.0165 | 0.018 | 0.0195 |

※ 使用する有意水準は、実際の情報時間（観察イベント数）で再度計算される

Janjigian et al. (2023)

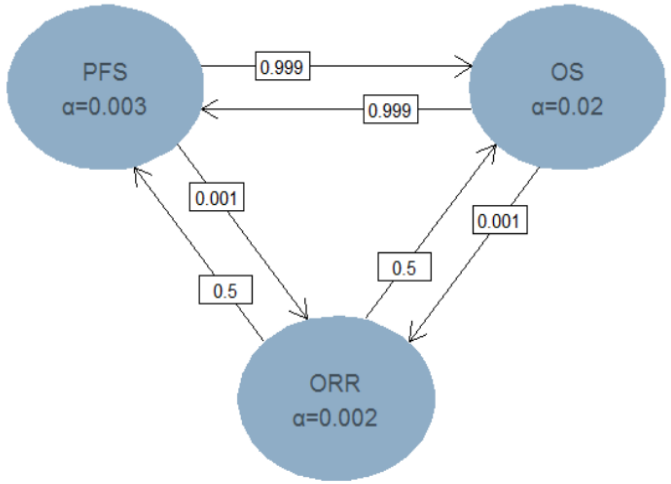
# 再掲：検定の多重性に関する統計記載

- 多重エンドポイントのα調整　グラフィカルアプローチを使用

An extension of the **graphical method** of Mauer and Bretz was used to control the overall type 1 error at a **one-sided α-level of 0·025**, with 0·002 originally allocated to objective response rate, 0·003 allocated to progression-free survival, and 0·02 to overall survival.

全体のα=0.025

Figure 4　Type I Error Reallocation Strategy




- 中間解析と最終解析間のα調整

有意水準はLan-DeMetsのα消費関数で更新

Statistical boundaries at each interim analysis were updated using a **Lan-DeMets O'Brien-Fleming spending fraction** on the basis of the information and α-level of the endpoint at each interim analysis.

Janjigian et al. (2023)

# Take home message

- がん臨床試験では**多重エンドポイント**、**多群比較**、**群逐次試験**など複数回の検定を行う状況において多重性の問題が生じる

- 検定の多重性の調整法は多岐にわたるが、CQを解決できる臨床的に解釈可能な方法を選択すべき（統計家だけで決定できない）

- **グラフィカルアプローチ**により仮説の構造や検定手順を明確にすることで臨床家と統計家の議論が円滑になることが期待される

# 参考文献


- 坂巻 顕太郎，寒水 孝司，濱﨑 俊光．統計解析スタンダード 多重比較法．朝倉書店（2019）
- Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. Statistics in Medicine 2009; 28: 586-604.
- Food and Drug Administration (FDA). Guidance for Industry: Multiple Endpoints in Clinical Trials. (2022).
- Holm S. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 1979; 6: 65-70.
- Janjigian YY, Kawazoe A, Bai Y et al. Pembrolizumab in HER2-positive gastric cancer. The New England Journal of Medicine 2024; 391: 1360-1362.
- Janjigian YY, Kawazoe A, Bai Y et al. Pembrolizumab plus trastuzumab and chemotherapy for HER2-positive gastric or gastro-oesophageal junction adenocarcinoma: interim analyses from the phase 3 KEYNOTE-811 randomised placebo-controlled trial. The Lancet 2023; 402: 2197-2208.
- Janjigian YY, Kawazoe A, Yañez P et al. The KEYNOTE-811 trial of dual PD-1 and HER2 blockade in HER2-positive gastric cancer. Nature 2021; 600: 727-730.
- Kato K, Machida R, Ito Y et al. Doublet chemotherapy, triplet chemotherapy, or doublet chemotherapy combined with radiotherapy as neoadjuvant treatment for locally advanced oesophageal cancer (JCOG1109 NExT): a randomised, controlled, open-label, phase 3 trial. The Lancet 2024; 404: 55-66.
- Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika 1983; 70: 659-663.
- Maurer W. Multiple comparisons in drug clinical trials and preclinical assays: a-priori ordered hypothesis. Biometrie in der Chemisch-pharmazeutischen Industrie 1995; 6: 3-18.
- Maurer W, Bretz F. Multiple testing in group sequential trials using graphical approaches. Statistics in Biopharmaceutical Research 2013; 5: 311-320.
- Wiens BL. A fixed sequence Bonferroni procedure for testing multiple endpoints. Pharmaceutical Statistics 2003; 2: 211-215.
- Wiens BL, Dmitrienko A. The fallback procedure for evaluating a single family of hypotheses. Journal of Biopharmaceutical Statistics 2005; 15: 929-942.