

中間解析の基礎と実際の運用

国立がん研究センター 生物統計部

JCOGデータセンター 統計部門

水澤 純基

2018.3.14

注：わかりやすさを優先しているため、統計的には正確ではない表現も含まれます

営利目的でのご利用や無断転載はご遠慮ください

ICR臨床研究入門 <http://www.icrweb.jp>

中間解析 (interim analysis) とは？

- あらかじめプロトコールに規定された方法に従って、試験終了前に有効性または安全性に関して意思決定が可能な結論が得られたかどうかを判断する目的で行うすべての解析
 - キーオープン（真の割付群で群間比較）して実施する
 - 盲検下レビューや進捗管理（モニタリング）は中間解析ではない

なぜ中間解析は必要か？

- 中間解析結果に基づいて登録中止・結果の早期公表を行い、**不利益を被る患者や回避可能な不利益の発生を最小化する**
 - 試験治療が予想よりも良い成績の場合、速やかに結果公表（**有効中止**）
 - これから治療を受ける将来の患者さんのため
 - 試験治療が予想よりも悪い成績/試験治療の有効性を示す可能性が低いことがわかった場合、速やかに結果公表（**無効中止**）
 - 登録中の中間解析の場合、これから試験に登録される患者のため
 - 臨床で用いられている治療法の場合、治療を受ける可能性のある患者を最小限に

ヘルシンキ宣言

「8. 医学研究の主な目的は新しい知識を得ることであるが、この目標は個々の被験者の権利および利益に優先することがあってはならない。」

中間解析時の統計的注意事項：検定の多重性

- 何度も検定を行うと、 α エラーを起こす確率が事前に設定した有意水準(5%)を超えてしまう問題のこと
 - 「 $\alpha=5\%$ と設定する」 = 「差がないのに差があると言ってしまう確率が5%以下」を満たさなくなることが問題

比較の回数	どこかで「有意差 $p < 0.05$ 」が出る確率 (%)
1	5.0
2	9.7
3	14.3
4	18.5
5	22.6
10	40.1
20	64.2

10個サブグループ解析をしたら、本当は差がなくても1個くらいは有意になる確率が40%

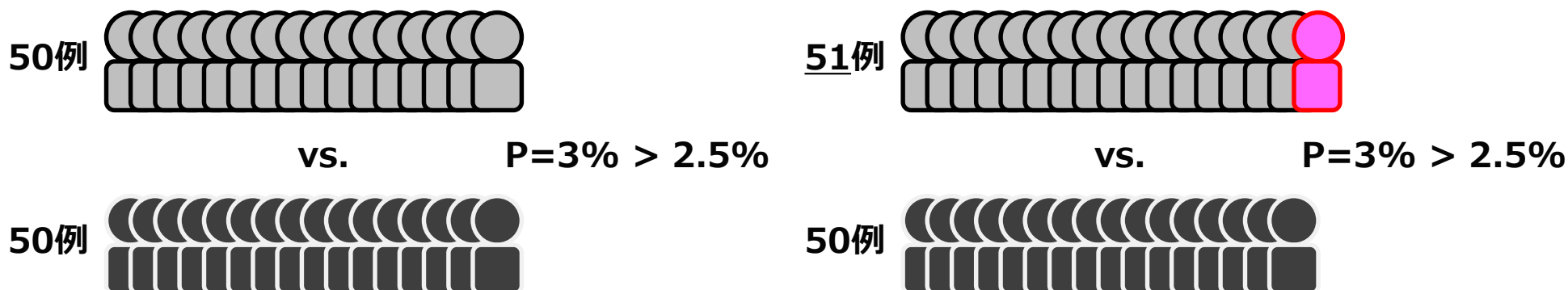
ボンフェローニの方法

- 検定回数(k)に応じて、1つ1つの検定の有意水準を $\alpha \div k$ で行う
 - $\alpha=5\%$ 、検定回数=2なら、1つの検定は $5\% \div 2=2.5\%$ で行う
 - $p < 2.5\%$ なら有意と判断
 - $\alpha=5\%$ 、検定回数=3なら、1つの検定は $5\% \div 3 \approx 1.67\%$ で行う
 - $p < 1.67\%$ なら有意と判断
- 合計が5%になるなら、1つ1つの検定の α は均等ではなくても良い
 - 1%と4%などにしてもOK
 - 例えば、70歳未満の集団では $\alpha=4\%$ 、70歳以上の集団では $\alpha=1\%$ で検定
 - 70歳未満で $p=3\%$ 、70歳以上で $p=2\%$ なら、70歳未満のみ有意と判断

中間解析で生じる検定の多重性の問題

1. 中間解析時のデータは最終解析にも含まれる

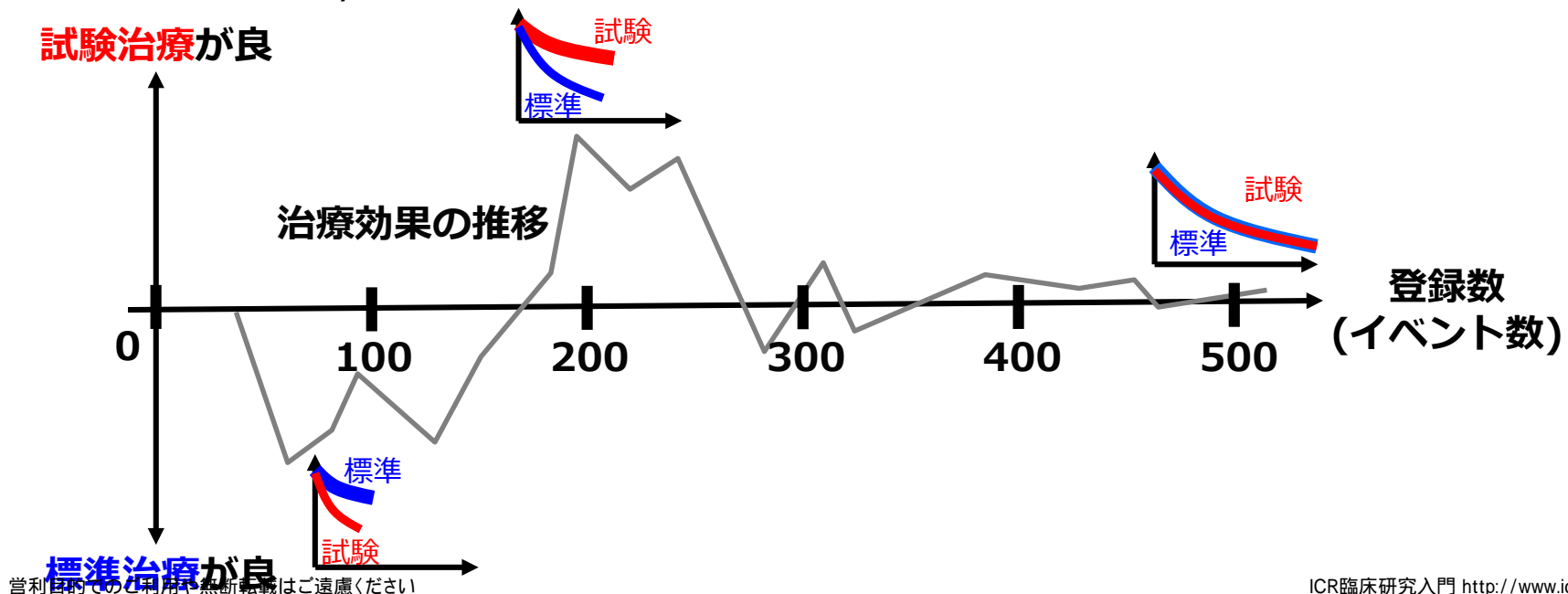
- 中間解析と最終解析のデータは相関する
- 中間解析時の登録数(イベント数)が100例、最終解析時の登録数(イベント数)101例の時、2回検定するため、ボンフェローニ法でそれぞれ2.5%ずつ検定？



- ボンフェローニの方法がベストとはいえない (αがもったいない)
- 統計的には、かなりサンプルサイズを増やさないと検出力が落ちる
 - 良い治療法を誤って捨ててしまう可能性が高くなる

中間解析で生じる検定の多重性の問題

2. 中間解析の回数やタイミングは計画段階通りにならないこともある
 - 効果安全性評価委員会から追加で中間解析するように言われることもある
 - データマネージメントなどの関係で予定通りのタイミングでの実施は困難
3. 試験の途中では、極端な結果が得られやすい
 - 試験の早期段階では相当大きな差がないと有効中止にしない方がよさそう
 - 登録数and/orイベント数が集まった後で結論が変わるリスクを避ける



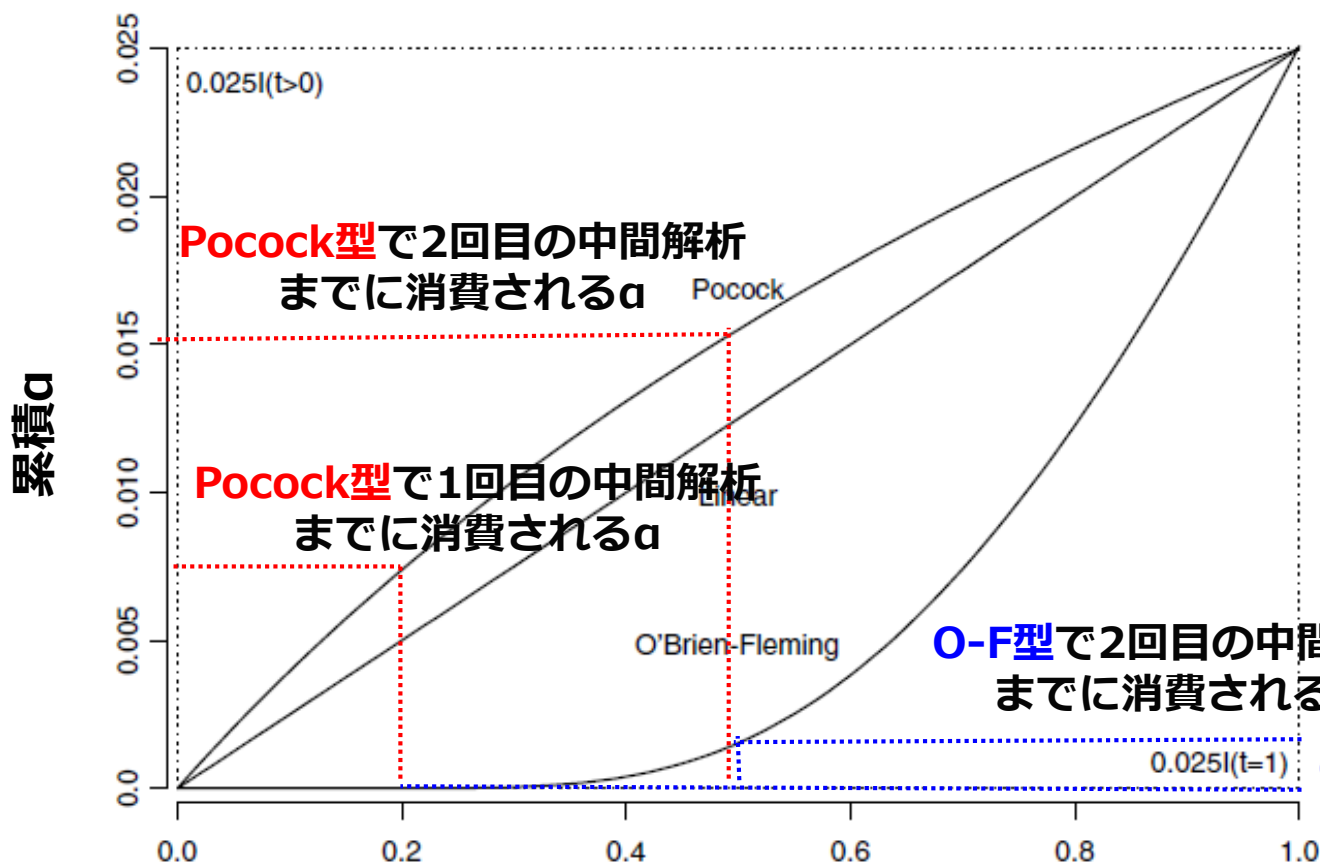
中間解析の α 調整の手法：

Lan & DeMetsの α 消費関数

- 先述の問題点を考慮した柔軟な手法
 - 解析の回数やタイミングが変更になっても α を調整できる
- 必要な情報（生存時間解析の場合。割合や連続値は患者数）
 - 最終解析時に予定しているイベント数
 - 解析時に観察されたイベント数
- 各解析時の α はinformation time[fraction]（情報時間）に比例
 - 解析時点でのイベント数÷最終解析時に予定しているイベント数
 - 中間解析時・最終解析時のイベント数20・100なら、情報時間=0.2
- α 消費関数に用いる関数によって各解析時点での α が変わる
 - **Pocock型**：どの解析時点でもほぼ同じ α
 - **O'Brien-Fleming (O-F) 型**：前半は小さい α 、後半は5%に近い α

型別の α 消費関数の図

- 情報時間が20%、50%、100%で解析し、 $\alpha=2.5\%$ とした場合



情報時間	Pocock型	O-F型
0.2	0.74%	<0.01%
0.5	0.98%	0.15%
1.0	1.30%	2.45%

それぞれの解析で、同じ対象を含んでいる（互いに相関がある）ので、合計して2.5%以上になる（ボンフェローニの方法よりお得）

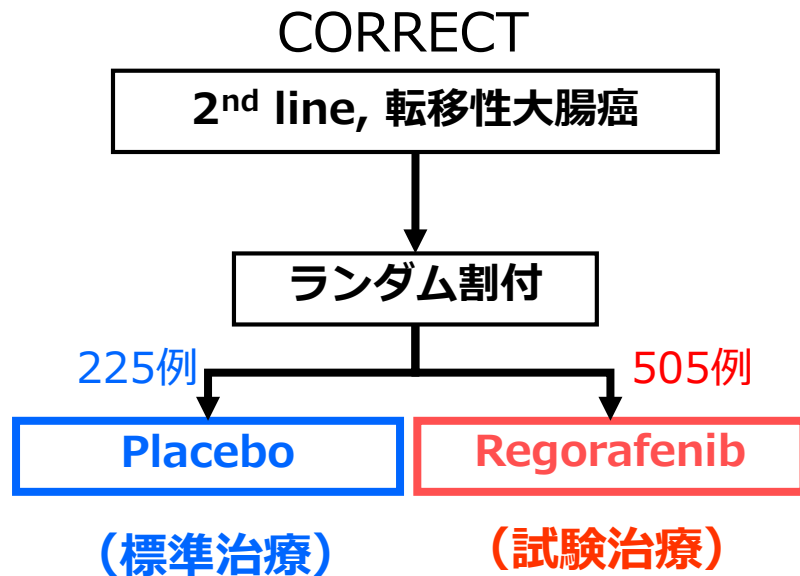
O-F型で2回目の中間解析までに消費される α
O-F型で1回目の中間解析までに消費される α

情報時間（観察イベント数÷最終解析時の期待イベント数）

有効中止・無効中止の原則

- 有効中止は統計学的規準に基づいて決めるのが標準
 - 統計学的に試験全体で α エラーを5%以下に制御
 - **O-F型**のLan & DeMetsの α 消費関数を用いることが多い
 - 上記の方法で算出された α よりも中間解析時のp値が小さければ有効中止
- 無効中止は厳密な統計学的規準を求めないことも多い
 - ベイズ流予測確率、条件付き検出力などの指標を用いて総合的に判断
 - JCOG試験の場合は点推定値で劣っていたら止めることも多い
 - 勝てる見込みがなさそうなら止める
 - 点推定値で劣る場合に試験治療を継続するメリットがほとんどないと考える
 - 事前に試験実施側とDSMCで詳細な打ち合わせが必要
 - 無効中止のためにどのくらいの効果・期間が必要か
 - 無効中止のために有効中止と同じくらいのエビデンスが必要か（有意に劣らないと止めない?）
 - 両群に差がない場合、予定通り試験を継続したいか?

中間解析で有効中止の例



Primary endpoint : OS
片側 $\alpha = 2.5\%$
Power = 90%
N(event) = 690(582)

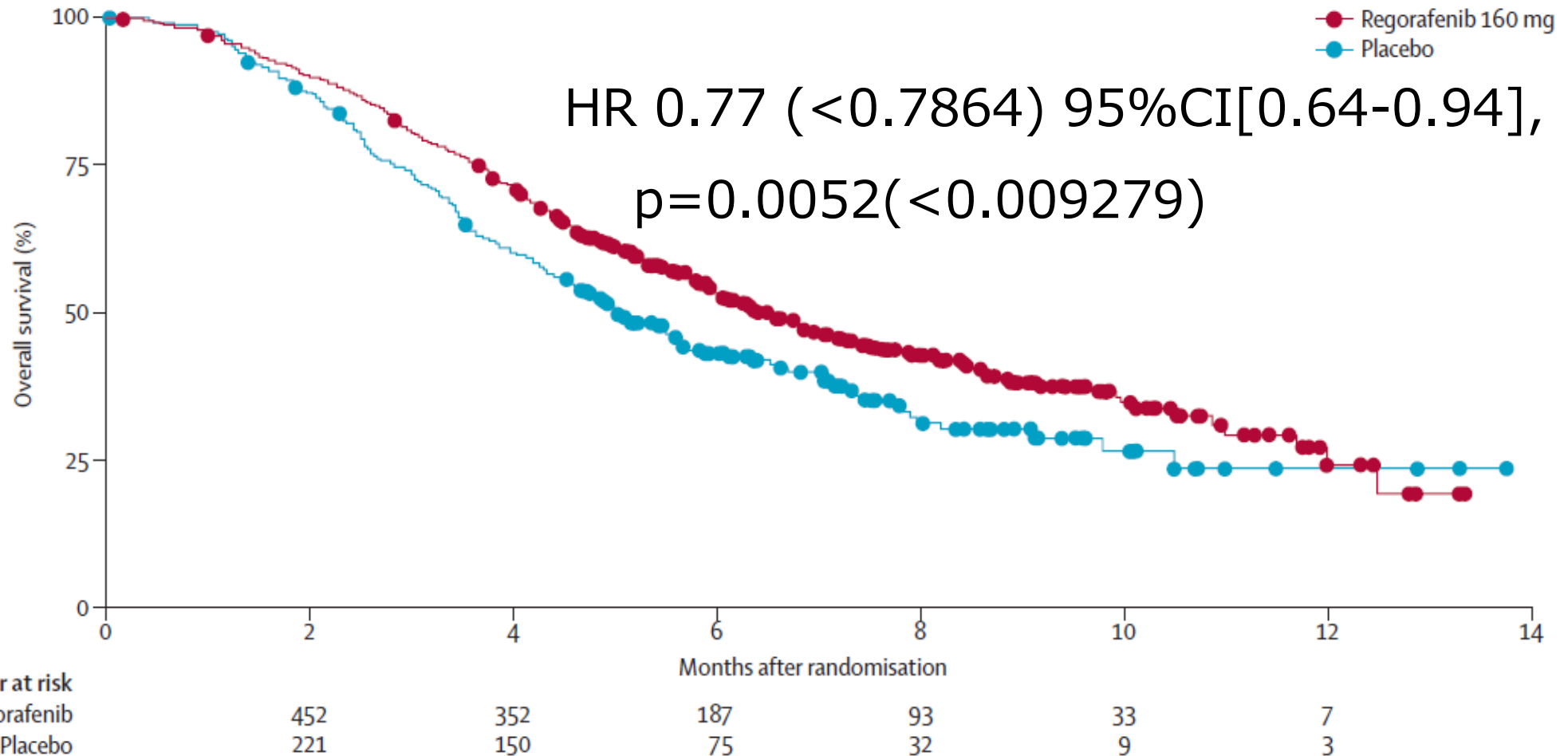
論文の統計記載の抜粋

- A **Lan-Demets alpha spending function** determined the monitoring boundary for efficacy so the overall false positive rate (α) was less than or equal to 0.025 (one-sided).
- The alpha spending function was the **O'Brien-Fleming type boundary** specified. Boundaries were specified to stop the study for efficacy or futility on the basis of the actual number of events included in the analysis.
- At the second interim analysis, the study was to be stopped for futility if the HR (regorafenib over placebo) was 0.9006 or greater, and for efficacy if the one-sided p value was less than or equal to 0.009279, roughly corresponding to an HR (regorafenib over placebo) of less than or equal to 0.7864.

- **O'Brien-Fleming型**の α 消費関数を使って有意水準を決定
- 2回目の中間解析で、 $p < 0.009279$ なら有効中止

事前に定めた有意水準 α をp値が下回り有効中止

A



中間解析で無効中止の例

NCCTG N1047

Stage III colon cancer
Wild-type KRAS

ランダム割付

909例

954例

mFOLFOX6

mFOLFOX6 + Cet

(標準治療)

(試験治療)

Primary endpoint : DFS

両側 $\alpha = 5\%$

Power = 90%

N(event) = 2070(515)

論文の統計記載の抜粋

- The [decision to terminate enrollment](#) was made by the study team following a recommendation from an external [data and safety monitoring committee \(DSMC\) review of a nonprespecified conditional power](#) analysis.
- At the time of the second planned interim analysis, the results indicated that [the experimental group would not be better than the control group](#).
- Based on this information, the authors informed the DSMC for the trial through a confidential presentation of the analysis.
- [The DSMC then voted to recommend closure of accrual to the trial](#), with the study team then being informed of this recommendation. With this recommendation, the study team was obligated to either halt accrual or to provide additional information to the DSMC to alter their decision.

• 無効中止の統計的な規準はプロトコールに記載されていない

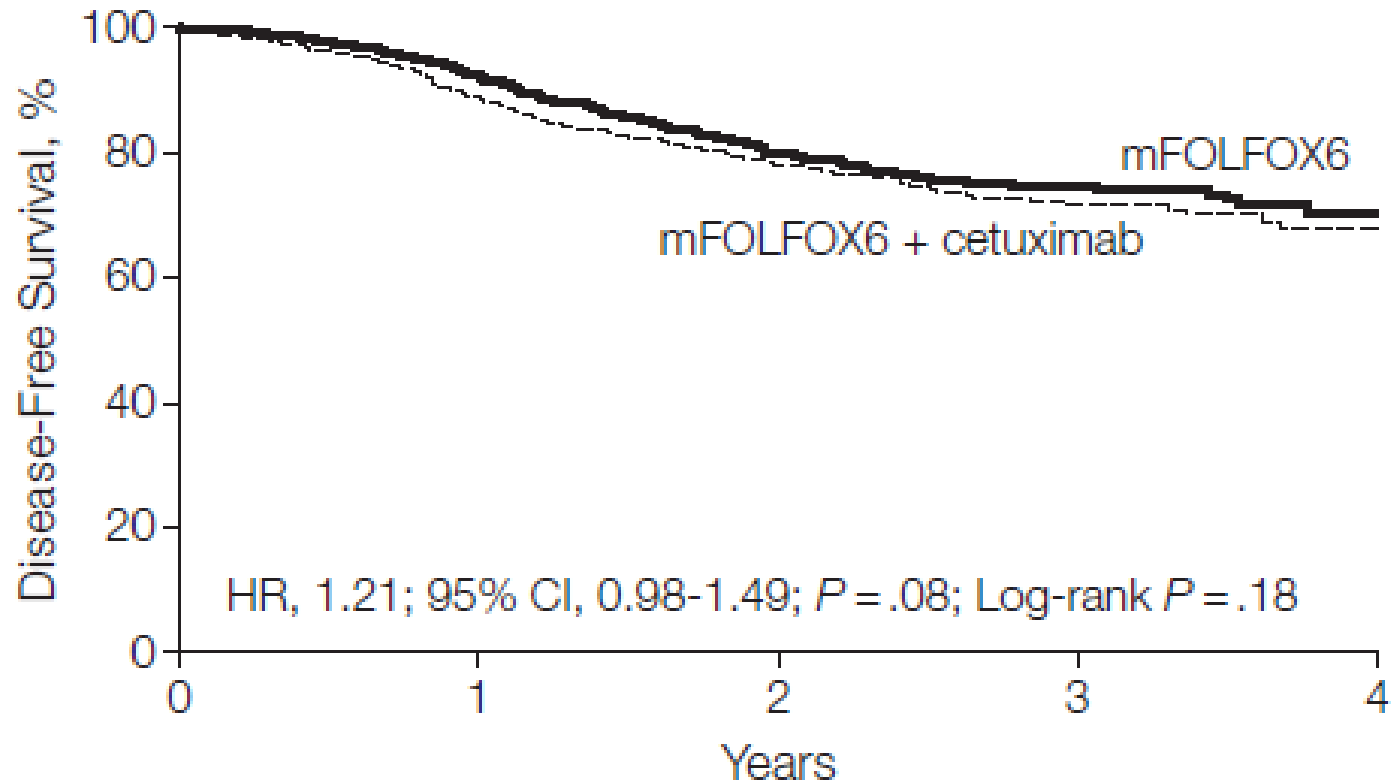
• Non-prespecifiedな解析結果に基づいてDSMCが無効中止を勧告

• Alberts SR et al. JAMA. 2012;307(13):1383-93.

• <http://www.jco.org/doi/full/10.1200/JCO.2012.46.2457>

第2回中間解析の結果：無効中止

- HR, 1.21 95%CI[0.98-1.49]; p=0.18
- 統計的に有意に劣っているわけではないが無効中止



No. at risk					
mFOLFOX6	909	659	413	163	48
mFOLFOX6 + cetuximab	954	667	417	154	39

Short summary

- 中間解析の目的
 - 有効性または安全性に関して意思決定が可能な結論が得られたかどうかを判断
 - 患者さんの不利益を最小限にする
- 中間解析の検定の多重性の問題
 - 何度も検定することで α エラーが増大する
 - 試験全体で α エラーを有意水準以下（ $\approx 5\%$ ）に抑える必要がある
- 多重性の調整のための手法
 - Lan & DeMetsの α 消費関数にO-F型を用いることが一般的
 - 情報（サンプルサイズやイベント数）が少ない時は小さい α を設定
 - 中間解析時の α よりもp値が小さければ有効中止
 - 無効中止は統計学的規準だけではなく、総合的な判断のこともある
 - 必ずしも検定で有意に劣る必要はない

中間解析審査の実際

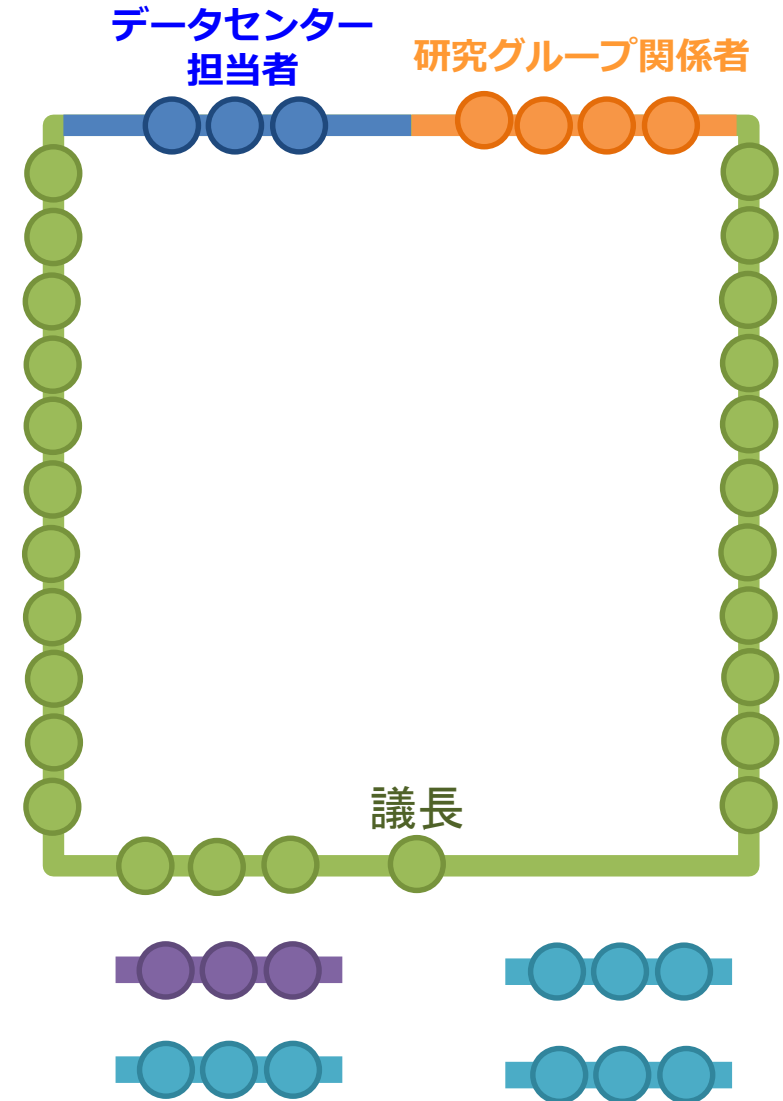
Japan Clinical Oncology Group (JCOG) の場合
詳細はJCOGポリシー「22 中間解析の方法と審査」を参照
http://www.jcog.jp/basic/policy/A_020_0010_22.pdf

誰が行うか？ -会議審査参加者の構成-

- ① 審査員（効果安全性評価委員）
- ② 研究グループ関係者（研究者）
- ③ JCOGデータセンター担当者
- ④ 効果安全性評価委員会(DSMC)事務局
- ⑤ オブザーバー

- 守秘義務 -

すべての参加者は秘密保持に関する書類に署名



誰が中間解析結果（有効性の群間比較）を見られる？

見られる

当該試験と直接的な関係がなく、早期の結果を知っても試験にバイアスが生じる可能性は低いと考えられる者

- ① 審査員（DSMC）
- ③ データセンター担当者のうち、中間解析担当統計家とデータセンター長
- ④ DSMC事務局

見られない

早期の結果を知ることによって、試験にバイアスが生じる可能性がある者

- ② 研究グループ関係者（研究者）
 - 片群の結果が良い/悪いことを知って、担当医が登録に消極的になる
 - 登録が完遂できず、結論が得られなくなる
- ③ データセンター担当者のうち、試験責任統計家
 - デザインや解析法の変更により、 α エラーの増大

どのように行うか？-オープンセッション-

① 研究の説明と質疑応答（オープンセッション）

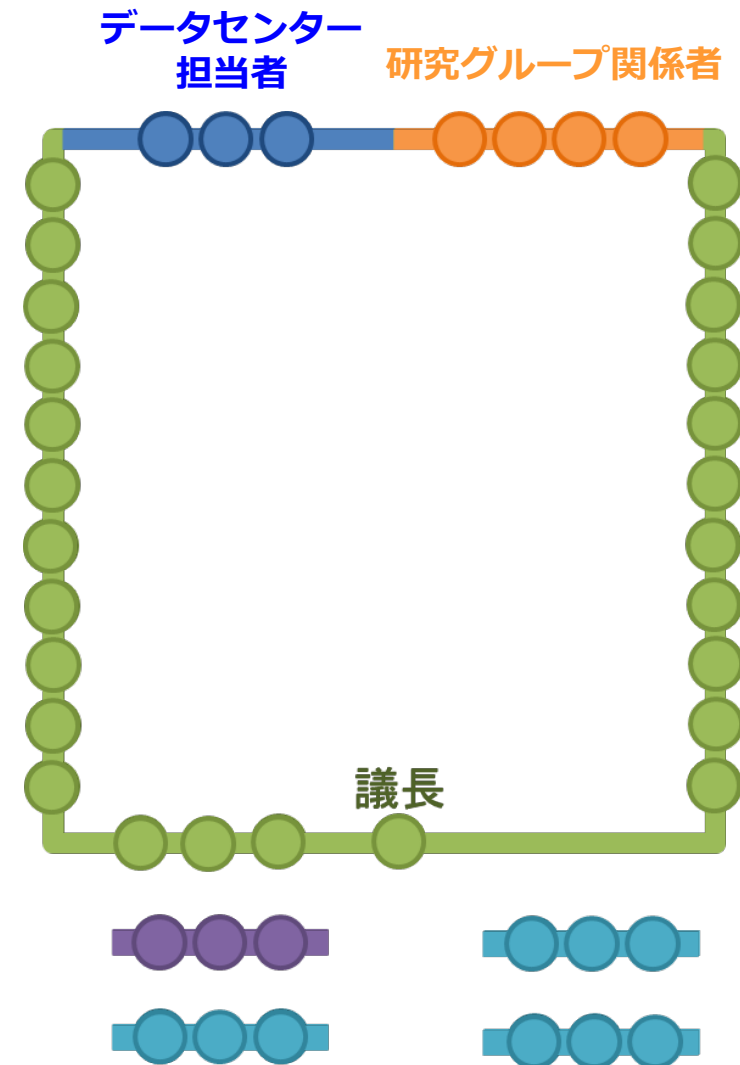
- 研究事務局/研究グループの代表者
 - 試験の概要・進捗、安全性データ等を説明
- グループ担当の統計家（試験責任統計家）
 - 試験デザイン、解析法、意思決定ルール等を説明
- 質疑応答
 - どの程度の結果で無効中止しても良いか
 - 現状の日常臨床の状況
 - 今後の治療開発の方向性



当該研究関係者はすべて退室

- 研究グループ関係者
- グループ担当の統計家
- 効安委員のうちグループに属する研究者など

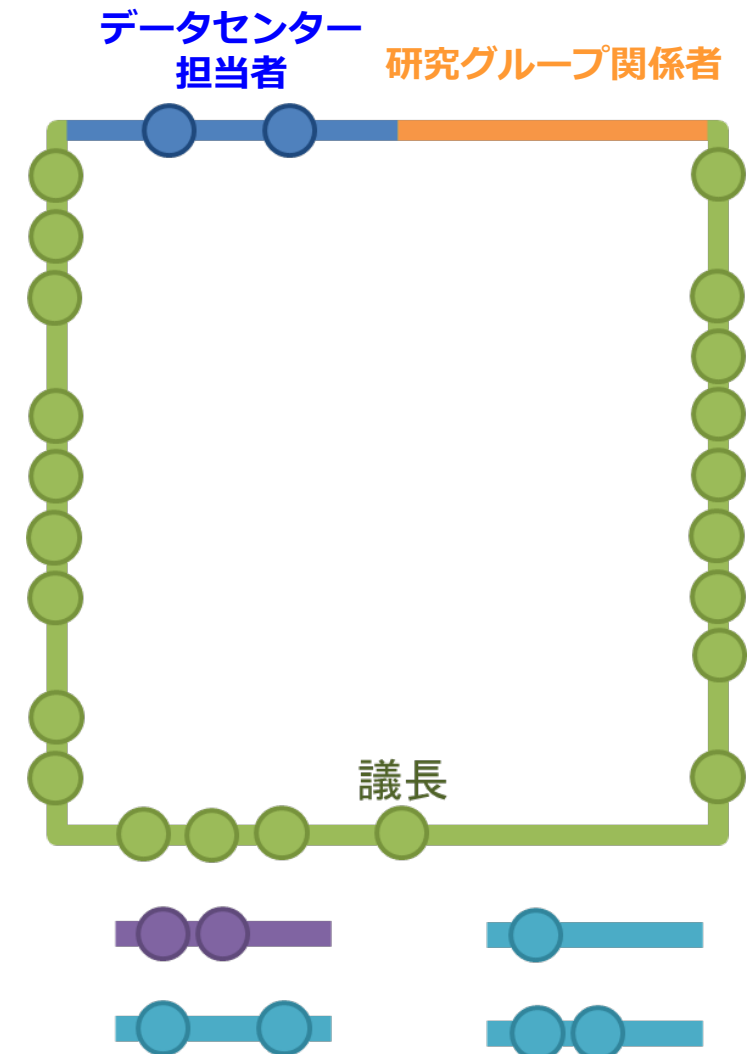
営利目的でのご利用や無断転載はご遠慮ください



どのように行うか？-クローズドセッション-

② 審議（クローズドセッション）

- 中間解析レポートを配付
- 中間解析担当統計家が結果を説明
 - p値、治療効果の大きさ、
中止規準を満たしているかどうか
- 意思決定ルール等を考慮して議論
 - 特に、事前に定めた規準を超えて中止する場合は慎重な議論が必要
 - 無効中止するかどうかは、オープンセッションでの質疑応答が判断材料に

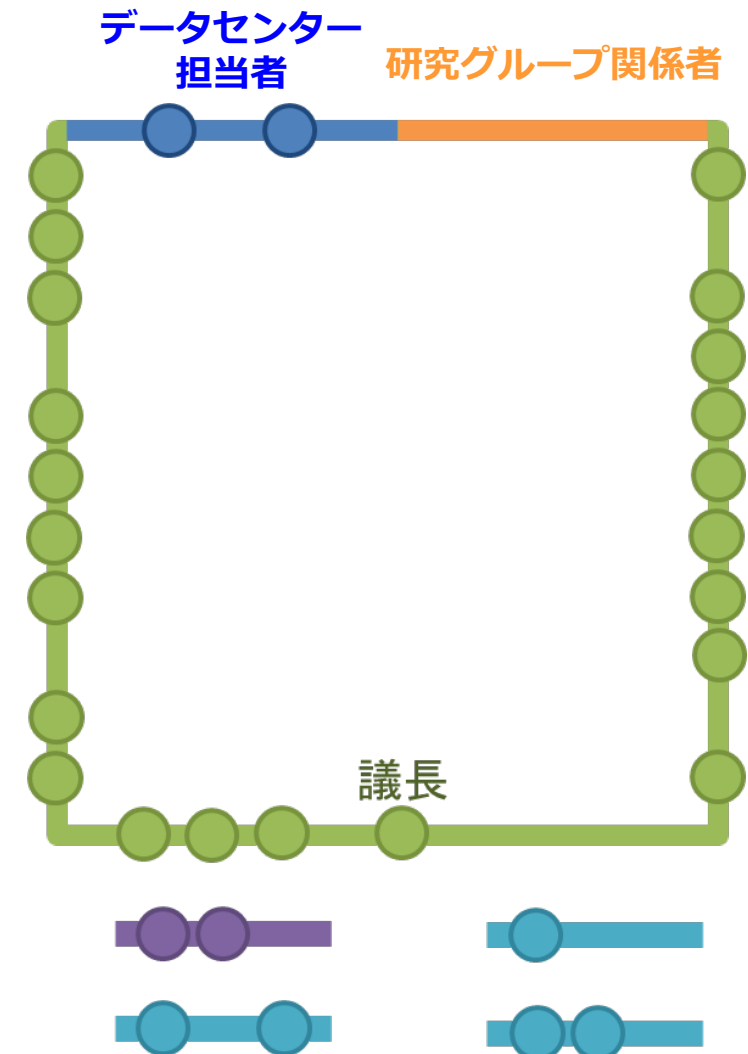


どのように行うか？-クローズドセッション-

③ 議決（クローズドセッション）

- 採決の選択肢
 - 研究の継続を承認
 - 研究の中止を勧告
 - 保留
- プロトコールの規定に従った判断
 - 過半数の賛同で可決
- プロトコールで規定されていない判断
 - 3分の2以上の賛同で可決

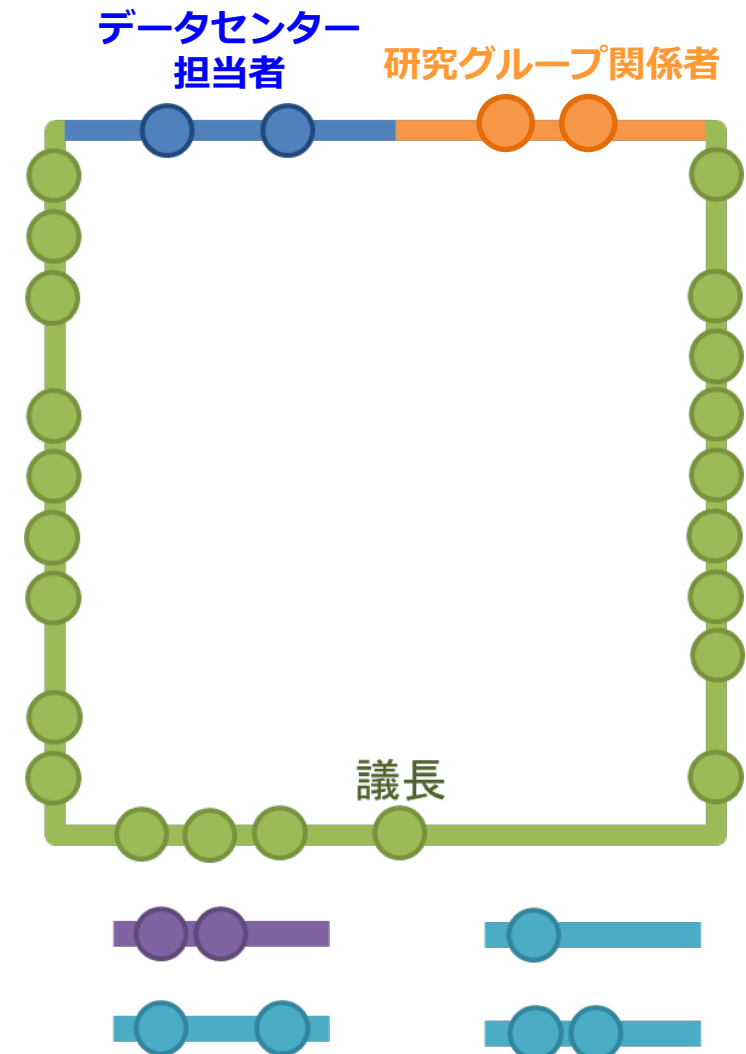
少なくとも、DSMCのSOPなどに、
議決の方法を明記しておく必要がある



どのように行うか？-クローズドセッション-

③ 議決の続き（クローズドセッション）

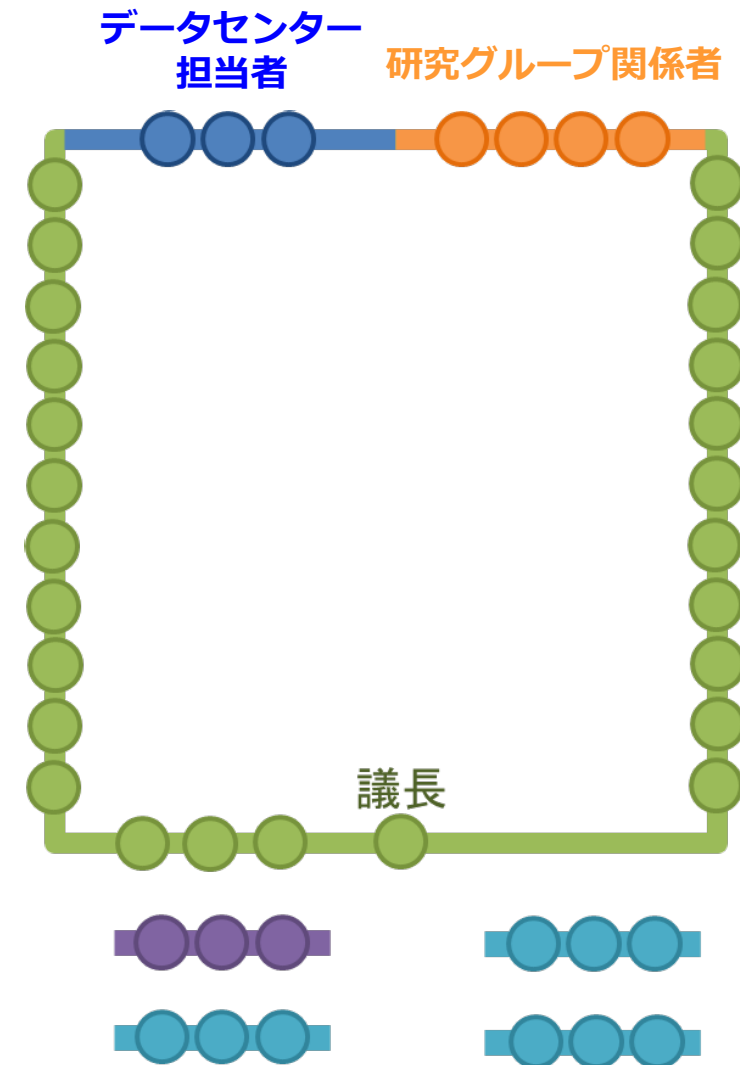
- 採決結果が以下の場合
 - 「研究の中止を勧告」
 - 「予め表明されていた意向に反する勧告」
- 最終決定前にグループから代表2名を入室させ、意見を聴取
(必要に応じて中間解析レポートを開示)
 - 効安の考えを説明する
- グループ関係者を再び退室させる
- 最終採決を行う
- 中間解析レポートを回収
(中止勧告の場合を除く)



どのように行うか？-オープンセッション-

④ 審査結果の通知（オープンセッション）

- 退室した当該研究関係者が入室
- 審査結果を通知
 - 継続の可否
 - 結果公表の可否
 - プロトコール改訂の要否…等
- 審査終了
- 後日、効安事務局、中間解析担当統計家、議長が議事録を作成



Take home message

- 中間解析の目的
 - 意思決定が可能な結論が得られたかどうかを判断
 - 患者さんの不利益を最小限にする
- 中間解析の検定の多重性の問題
 - 何度も検定することで α エラーが増大する
 - Lan & DeMetsの α 消費関数にO-F型を用いて試験全体の α を制御
- 中間解析の審査
 - 試験に直接関わらない審査員や事務局が必要
 - 試験に関わる研究者などは結果を見ることができない
 - 事前にSOPなどに、議決の方法や手順を定めておく
 - 研究者側の考えをよく理解し、中止の判断は慎重に