今さら聞けないがんを対象とした単群phase II試験

国立がん研究センター研究支援センター
生物統計部
水澤 純基

2021. 2.22

生物統計セミナー発展編

営利目的でのご利用や無断転載はご遠慮ください

典型的なphase2のデザイン

- 対象：がん種を限定
- 登録数：20-60例
- デザイン：**単群**でオープンラベル
- Endpoint：RECISTによる**奏効割合**
  - できるだけ早くphase III に進むかどうかを決める
    - 開発を続けるに値するかどうかの判断に用いる
    - 治療の利益というより、生物学的な**抗腫瘍活性**の存在を確認
  - **有効性**によるスクリーニング

ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

閾値と期待値

営利目的でのご利用や無断転載はご遠慮ください

第II相試験の意思決定に必要な知識：閾値と期待値

- **閾値と期待値**
  - **閾値**：これ以下の値であれば開発をあきらめる値
    - 最低限クリアして欲しい値
    - いわゆる「帰無仮説（≒意味のある効果なし）」に対応
  - **期待値**：これ以上の値であれば是非開発を続けたいと考える値

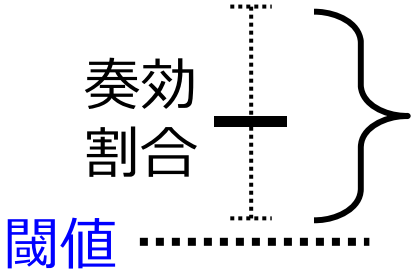
- どう決める？
  - 検定で意思決定する場合に直接関係するのは**閾値**
  - 「比べたい相手」＝「標準治療」は何かを考える
  - 標準治療と比較したリスクバランスを考える

ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

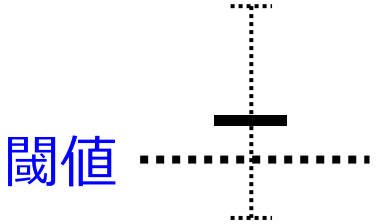
単群のphase 2試験はどうなったらpositiveか？

- 有意水準αに対応する割合の信頼区間が**閾値**を上回ればpositive

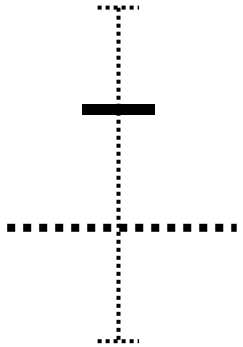


有意水準αに対応する
信頼区間

信頼区間が閾値を
跨ぐとnegative



割合が高くても
Nが少ないと精度が
十分ではなくnegative



割合がそれほど高くなくても
Nが多いと精度が高く
positive

検定でpositive/negativeの判断には期待値は
関係ないので、閾値の設定が肝

ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

ケース：どの設定が妥当？

条件設定：

- Primary endpointは奏効割合のPhase 2
- 標準治療の奏効割合は20%。
- 今回の試験治療は標準治療に上乗せすることになるため、毒性が強くなる

問い：あなたであれば、どの閾値を選択しますか？

1. 閾値=15%
2. 閾値=20%
3. 閾値=25%

ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

Phase 2 単群試験の期待値の設定

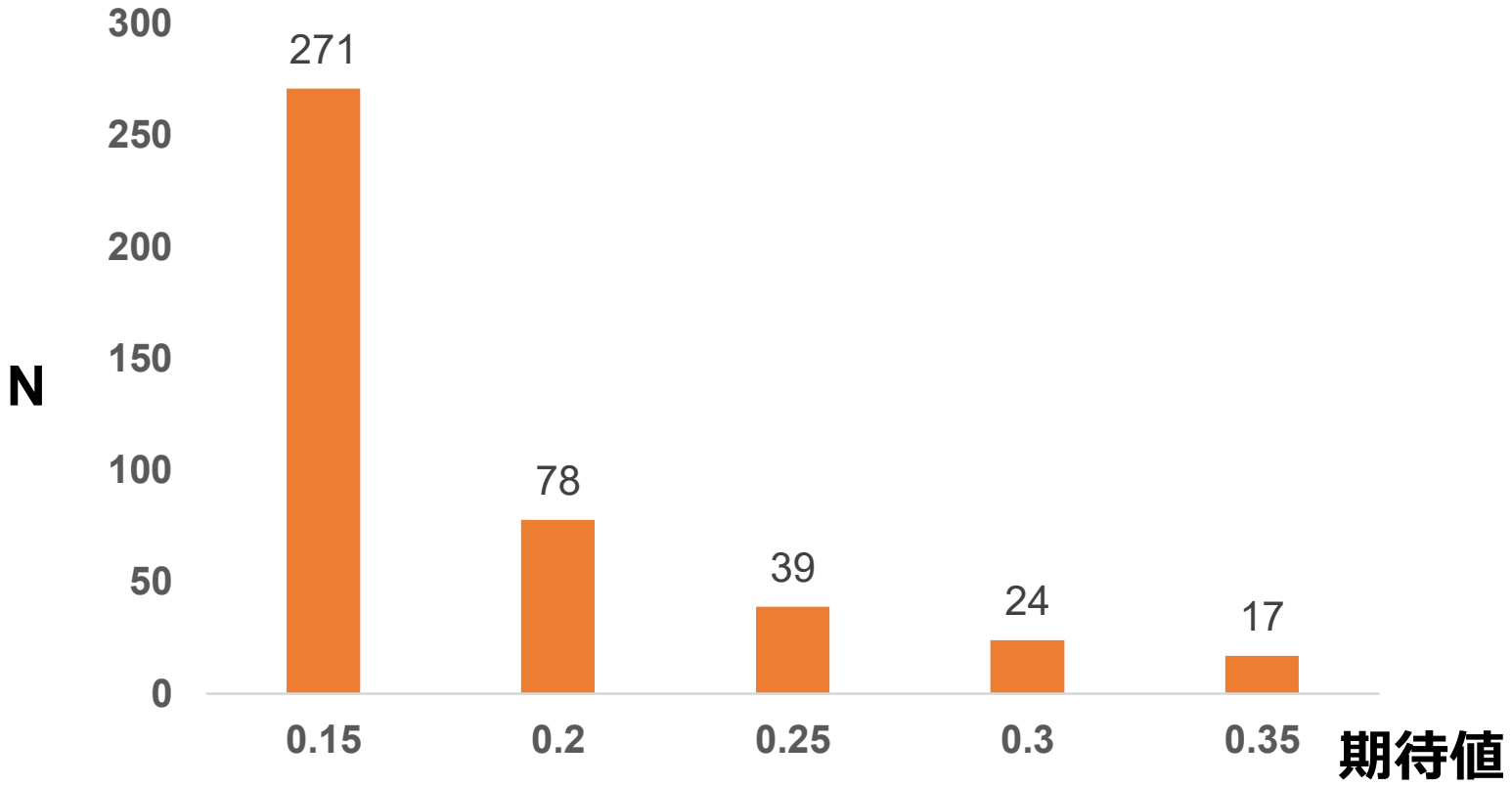
- 基本的な考え方は「そのくらいの値が得られるのであれば是非、開発を続けたい」と考える値
- 考慮すべきは以下の2点
  - 標準治療と比較した毒性とのバランス（臨床的意味）
  - 本当にそれだけの有効性が期待できるか
- 実際は、phase 2のサンプルサイズの目安はおおよそ決まっており、それに基づいて期待値を動かすことも多い
  - Nは少なすぎず、多すぎず、ほどほどに

営利目的でのご利用や無断転載はご遠慮ください

閾値・期待値とNの関係

- 大きな治療効果を期待するのであれば少ないNでOK

例：α=5%、検出力=80%、閾値=10%の場合



ICRweb 臨床研究入門 www.icrweb.jp

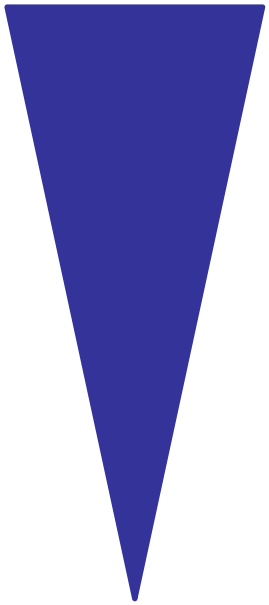
営利目的でのご利用や無断転載はご遠慮ください

解析対象数と信頼区間の幅の関係

解析対象数が増えれば、信頼区間の幅は狭くなる

割合の点推定値が10%の場合

- **10**人中1人　→ 95% CI[0.3%-44.5%]
- **20**人中2人　→ 95% CI[1.2%-31.7%]
- **50**人中5人　→ 95% CI[3.3%-21.8%]
- **100**人中10人　→ 95% CI[4.9%-17.6%]
- **200**人中20人　→ 95% CI[6.2%-15.0%]
- **1000**人中100人 → 95% CI[8.3%-12.0%]



ICRweb 臨床研究入門 www.icrweb.jp

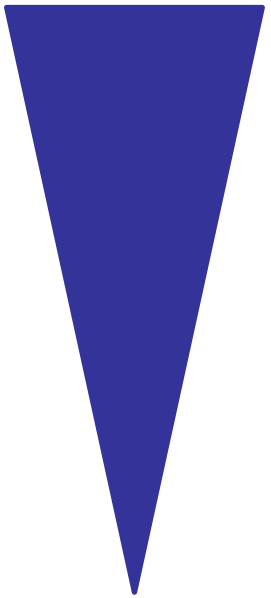
営利目的でのご利用や無断転載はご遠慮ください

信頼係数と信頼区間の幅の関係

信頼係数が小さくなれば、信頼区間の幅は狭くなる

割合の点推定値が10%の場合

- 50人中5人 → **99**% CI[2.2%-25.8%]
- 50人中5人 → **95**% CI[3.3%-21.8%]
- 50人中5人 → **90**% CI[4.0%-19.9%]
- 50人中5人 → **80**% CI[4.9%-17.8%]
- 50人中5人 → **50**% CI[6.8%-14.5%]



ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

検定と区間推定の関係

- 検定と区間推定は1:1の対応関係がある

片側有意水準（両側）	片側p値		両側p値		両側信頼区間(CI)
0.025（0.05）	P<0.025	=	P<0.05	=	95% CIが帰無仮説を含まない
0.05（0.1）	P<0.05	=	P<0.1	=	90% CIが帰無仮説を含まない
0.1（0.2）	P<0.1	=	P<0.2	=	80% CIが帰無仮説を含まない

P値をみなくても、設定した有意水準と信頼区間から
検定結果が有意かどうかがわかる

ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

理解のための問題：有意？ or 有意じゃない？

1. 閾値20%(=帰無仮説)、両側有意水準=5%の設定
   - 奏効割合 = 34% 95% CI[21.2%-48.8%]
2. 閾値20%、両側有意水準=5%の設定
   - 奏効割合 = 30% 95% CI[17.9%-44.6%]
3. 閾値20%、片側有意水準=5%の設定
   - 奏効割合 = 32% 90% CI[21.2%-44.5%]
4. 閾値20%、片側有意水準 = 5%の設定
   - 奏効割合 = 34% 95% CI[21.2%-48.8%]
5. 閾値20%、片側有意水準=5%の設定
   - 奏効割合 = 32% 95% CI[19.5%-46.7%]

営利目的でのご利用や無断転載はご遠慮ください

1-stage vs 2-stage デザイン

営利目的でのご利用や無断転載はご遠慮ください

単一標本デザイン(1段階法 or 1 stage法)

- 途中で解析・登録停止することなく、
  目標被験者数まで一気に登録
  - 最後に検定して有意（p値<α）ならpositive

- サンプルサイズ計算に必用な情報
  - 閾値(p0)
  - 期待値(p1)
  - α
  - β

ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

がんのphase II段階で真に有効なものは20%程度

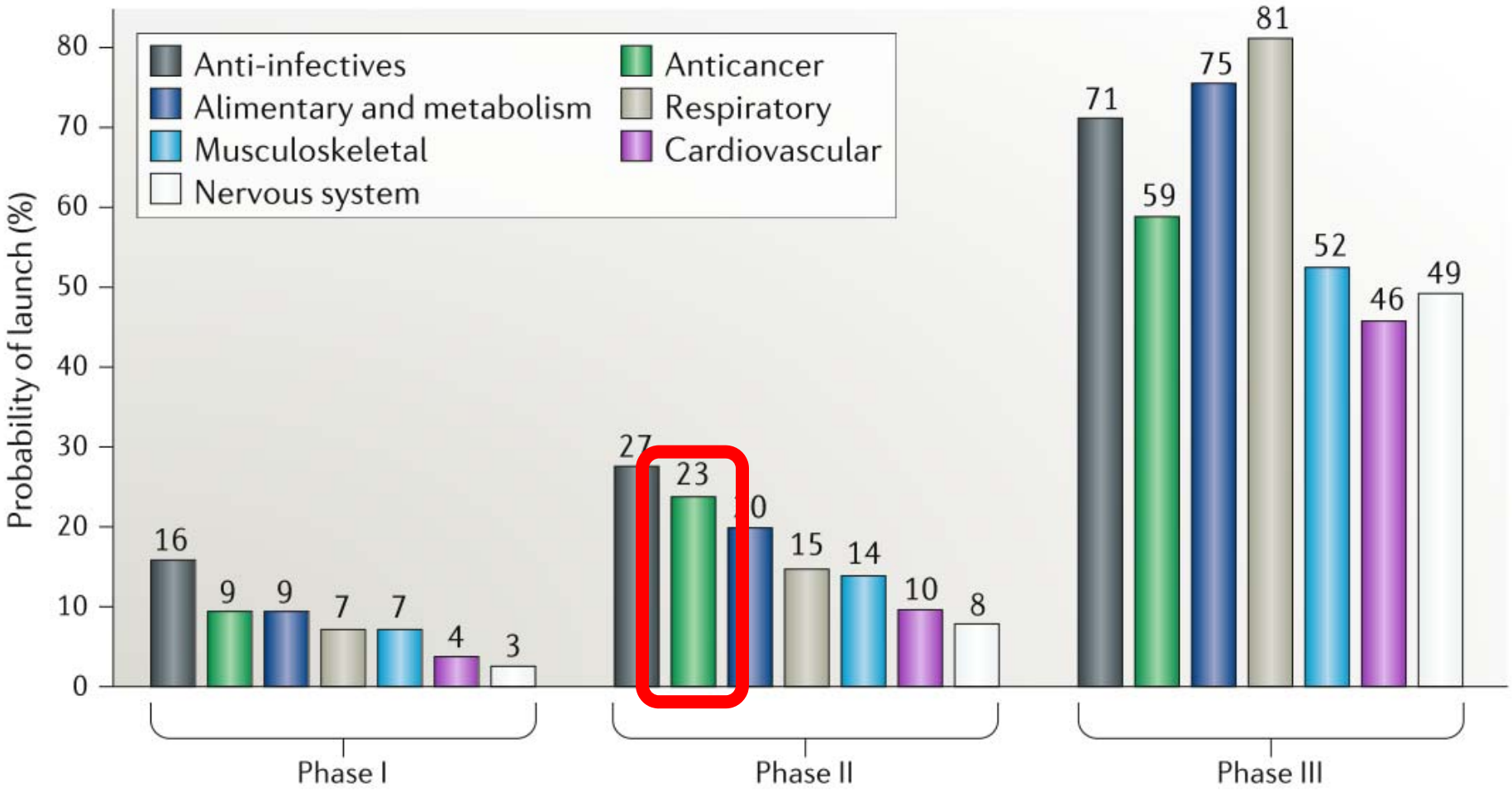


Fig. 2 | **Probability of launch from phases I, II and III by therapeutic area from 2010–2017.**

無効なものが多いのであれば、患者に対する倫理的配慮が必要

Dowden H, Munro J: Nat Rev Drug Discov 18:495-496, 2019

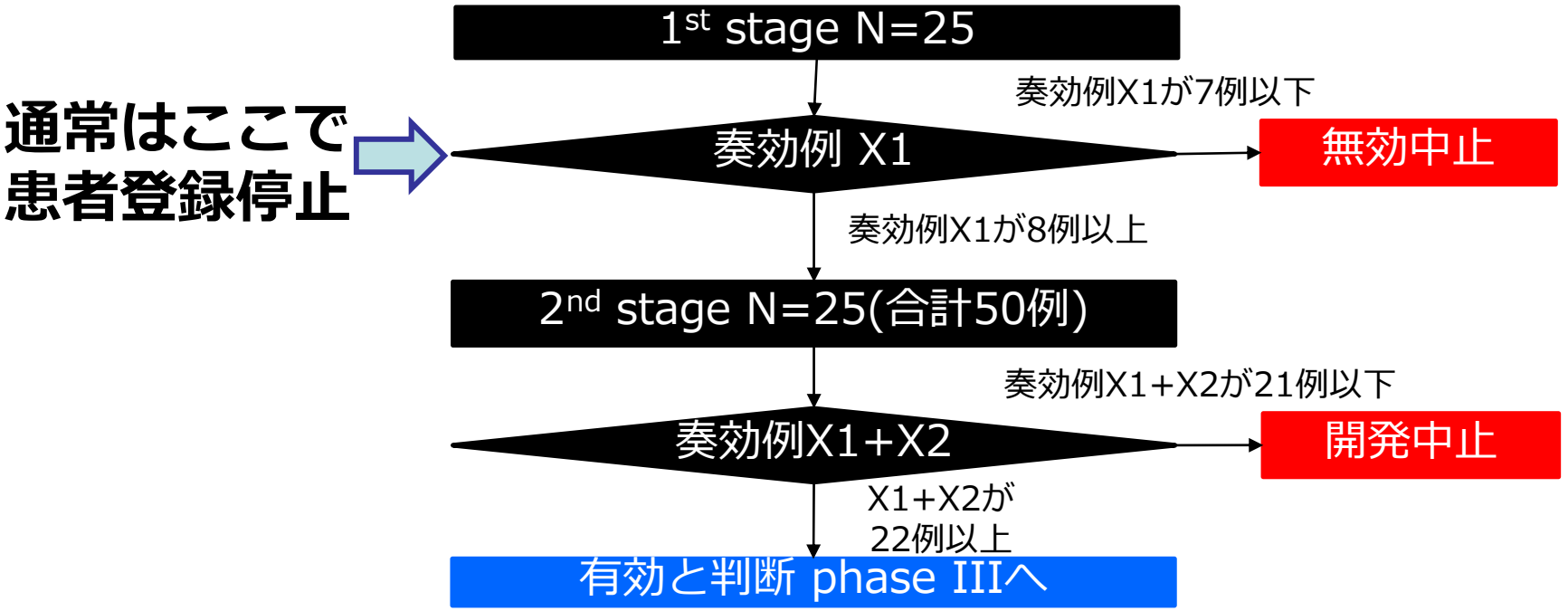
ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

2-stageデザイン

- 登録の途中で1度、有効性を評価する方法
  - 途中で中間評価を行う
  - 登録を止めて規準を満たすことが判明した後に登録再開が原則
- 例：JCOG0807（切除不能・再発食道がんに対するDCF療法）
  - 閾値35%(CF療法の奏効割合)、期待値50%、α=10%、検出力=80%

  → SWOGの2-stage design(後述)で50例。1st stageで25例

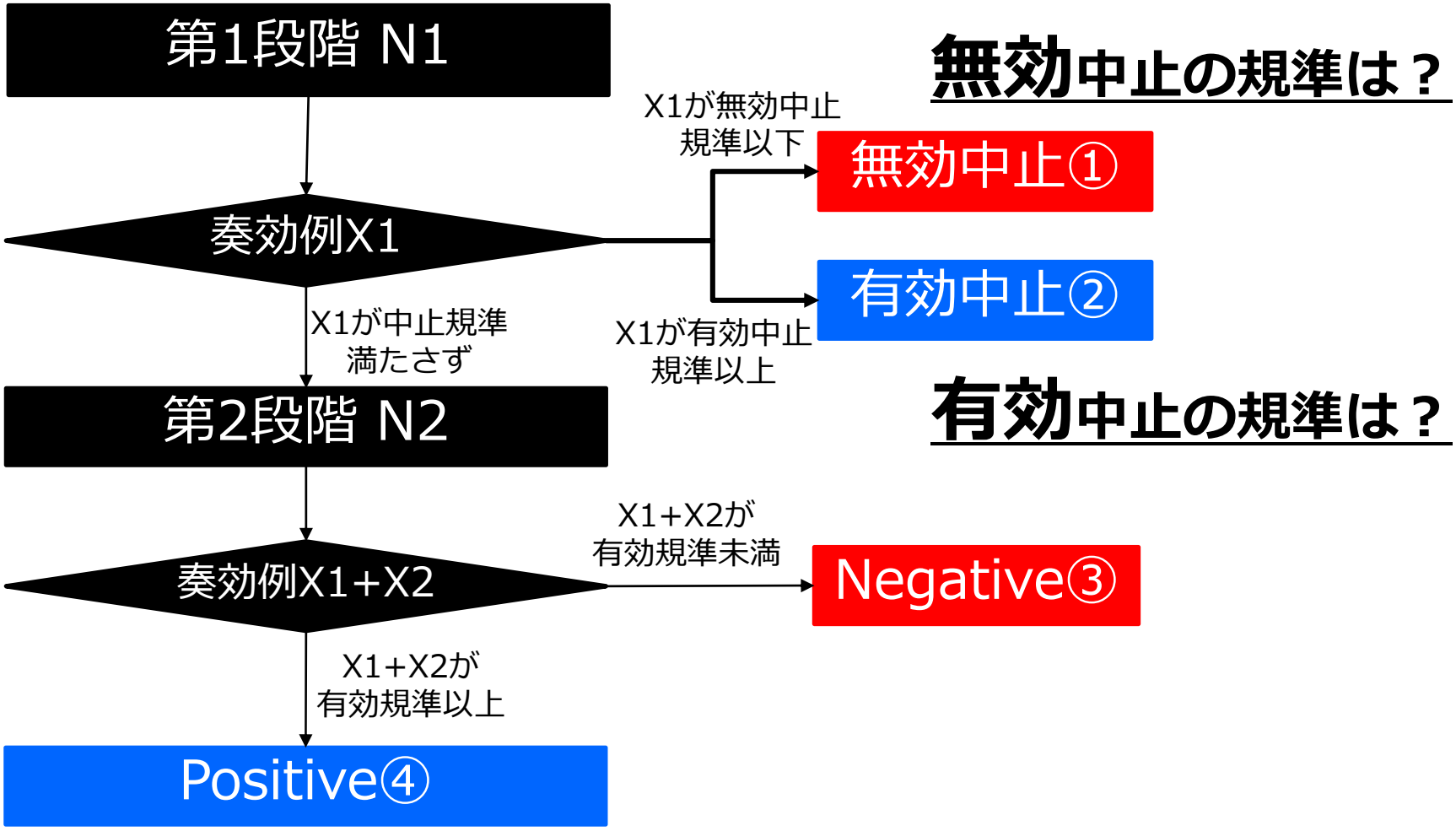


ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

2-stageデザインのバリエーション

無効中止と有効中止



ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

有効中止は必要か？

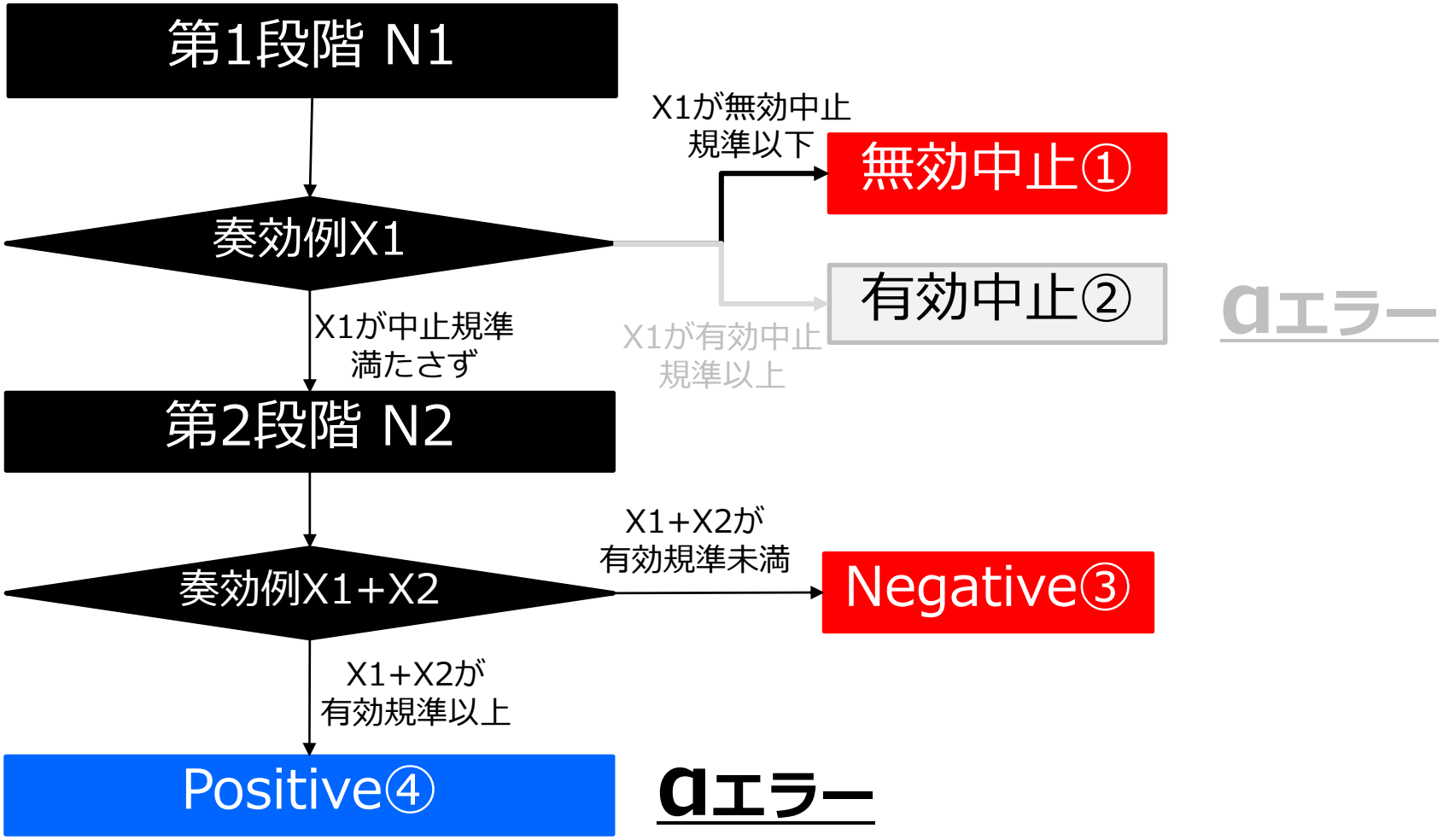
	Flemingの方法 1982	Simonの方法 1989	SWOGの方法 1992
無効中止の考慮	✓	✓	✓
有効中止の考慮	✓		

- 利用されている多くの2-stage designは無効中止のみ考慮
- その理由は？

営利目的でのご利用や無断転載はご遠慮ください

2-stageデザインの場合のαエラー

真実は効果がない場合

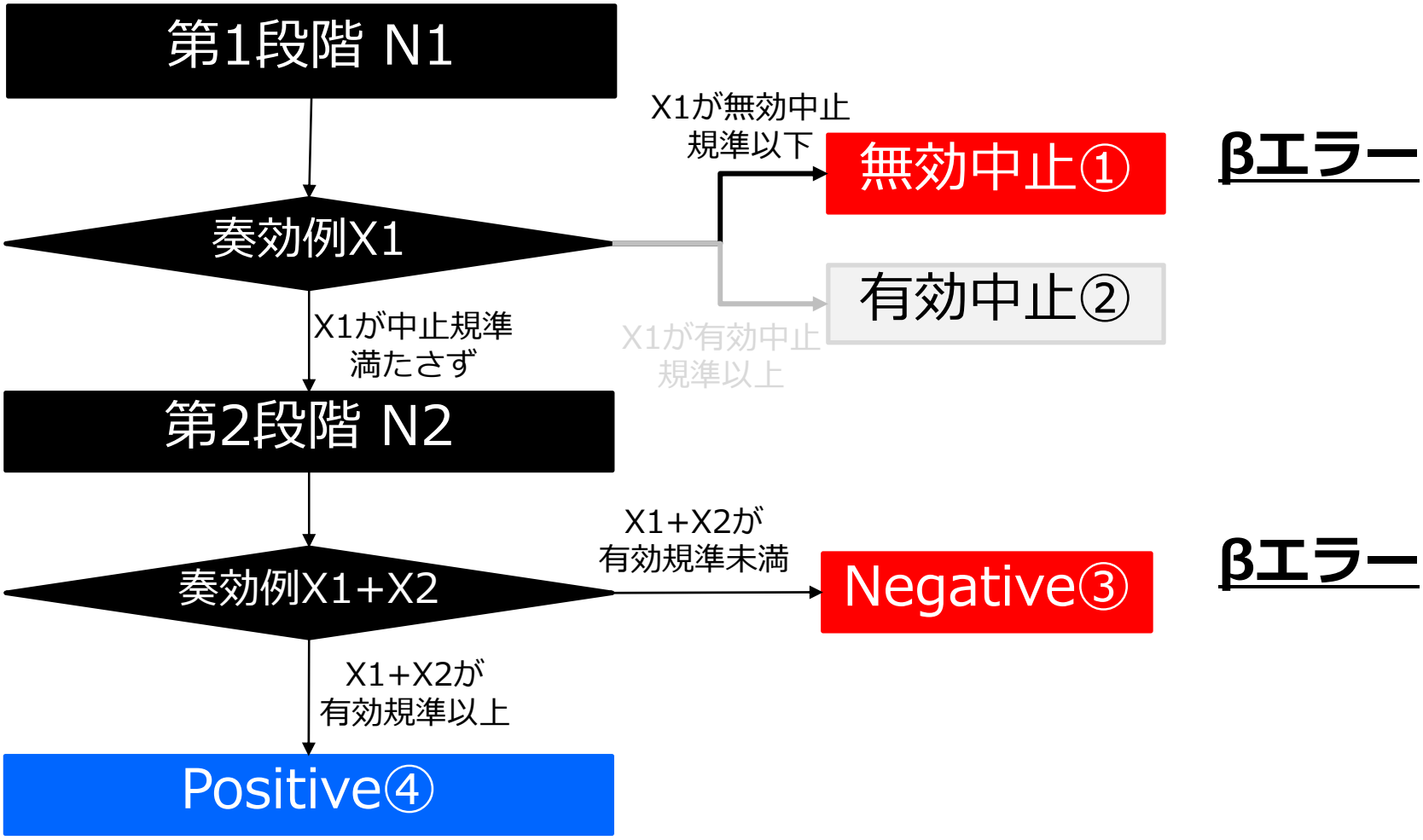


ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

2-stageデザインの場合のβエラー

真実は効果がある場合



ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

2-stageデザインのサンプルサイズの決め方

- αとβが条件を満たすサンプルサイズ（N）は複数パターンある
  - α=5%を固定しながら、β=20%を設定すると、
    （実際は条件付き確率でもう少し複雑だがイメージのため単純化）
    - 第1段階でβ=5%なら、第2段階はβ=15%
      - 第1段階でβが小さい = 第1段階で多くのNが必要
    - 第1段階でβ=15%なら、第2段階はβ=5%
      - 第1段階でβが大きい = 第1段階で相対的に少ないNで良い
- αとβが条件を満たすNの中から、何かしらの規準に従ってN(N1,N2)を設定する必要がある
  - 真の奏効割合がp0の時に期待されるNを最小にしたい
  - 真の奏効割合がp1の時に登録されるNを最小に
  - ・・・

ICRweb 臨床研究入門 www.icrweb.jp
営利目的でのご利用や無断転載はご遠慮ください

Simonの2-stageデザイン

- **無効中止のみ**考慮したデザイン
  - 一定以上の奏効例が得られているかどうかで有効/無効を判断
- サンプルサイズの決め方は2通りの規準を提案
  - **Optimal design**
    - **帰無仮説**が正しい時に**期待症例数**を最小にする
  - **Minimax design**
    - 特定の試験に登録される**最大症例数**を最小にする

Simon, Richard. *Controlled clinical trials* 10.1 (1989): 1-10.

ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

例：α=5%,検出力=80%

http://cancer.unc.edu/biostatistics/program/ivanova/SimonsTwoStageDesign.aspx
http://linus.nci.nih.gov/brb/samplesize/otsd.html

閾値=20%,期待値=40%	**Optimal**		**Minimax**
1st stageの患者数	13	<	18
奏効例がこれ以下なら無効中止	3 (23%)		4 (22%)
2nd stageまでの総患者数	**43**	>	**33**
奏効例がこれ以下なら無効	12 (28%)		10 (30%)
帰無仮説が真値の場合の期待患者数	**20.6**	<	**22.3**
帰無仮説が真値の場合に無効中止になる確率	75%		72%

閾値=5%,期待値=20%	**Optimal**		**Minimax**
1st stageの患者数	10	<	13
奏効例がこれ以下なら無効中止	0 (0%)		0 (0%)
2nd stageまでの総患者数	**29**	>	**27**
奏効例がこれ以下なら無効	3 (10%)		3 (11%)
帰無仮説が真値の場合の期待患者数	**17.6**	<	**19.8**
帰無仮説が真値の場合に無効中止になる確率	60%		51%

サンプルサイズは Optimal > Minimaxとなる

ICRweb 臨床研究入門 www.icrweb.jp
営利目的でのご利用や無断転載はご遠慮ください

Simonの2-stageデザインの欠点

- 第1段階,第2段階ともに患者が予定通りに登録される前提
- 多施設研究でそこまで正確に登録数を管理できるか？
  - 登録終了のアナウンスをしてもIC取得済みの患者は登録
  - 後から不適格例がわかったらどうする？
    - CRFが集まってから始めてわかる
    - 既に登録が終了していても、また登録再開？
- もし実際の解析対象数（全適格例）が変わると、設定しているα、βを保てなくなることがある
  - 無効中止、positiveの判断に必要な奏効例数はサンプルサイズ計算時の必要解析対象者数で規定
  - 別途、計算が必要となる

より現場のオペレーションに合ったデザインが必要

ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

SWOGの2-stageデザイン

- **無効中止**のみを考慮したデザイン
  - 第1段階はα=2%で無効中止の検討
    (期待値p1を帰無仮説とした検定)
    - おおよそ奏効割合がp0未満の時に無効中止になる
  - 第2段階はα=5.5%で検定
    - 有意ならPhase IIIへ
    - おおよそ奏効割合が(p0+p1)/2を超えると有意になる
- 利点
  - 実際の解析対象数が予定と比べて増減してもOK
    - 検定で判断する（p<αかどうか）ので柔軟に対応可

http://www.swogstat.org/stat/public/twostage.htm

Green, S et al., *Statistics in medicine* 11.7 (1992): 853-862.

ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

各方法のサンプルサイズと判断規準

	P0 vs P1	第1段階登録数	1段階目で無効	1段階目で有効	第2段階登録数	2段階目で無効	2段階目で有効	実質alpha	実施power
Simon	5% vs 20%	29	1	-	38	4	5	3.9%	90%
(Minimax)	10% vs 30%	22	2	-	33	6	7	4.1%	90%
	20% vs 40%	24	5	-	45	13	14	4.8%	90%
	30% vs 50%	24	7	-	53	21	22	4.7%	90%
SWOG	5% vs 20%	20	0	-	40	4	5	4.7%	92%
	10% vs 30%	20	1	-	35	7	8	2.0%	87%
	20% vs 40%	25	4	-	45	13	14	5.2%	91%
	30% vs 50%	30	8	-	55	22	23	4.1%	91%

営利目的でのご利用や無断転載はご遠慮ください

実際にどれを使うのか？

- 3段階以上のデザインは？
- 1段階と2段階の使い訳は？
- 2段階の中では？

営利目的でのご利用や無断転載はご遠慮ください

実際に多く使われているのは？：文献①

- 2010-2015年のJCO, Ann Oncol, BJC の phase 2 を調査
- Single armの315件中
  - 1-stage : 167件(53%)
  - 2-stage : 146件(46%)
    - Simonのoptimal : 47
    - Simonのmini-max : 21
    - Simonのoptimal or mini-max : 33
    - SWOG : 4
    - Fleming : 1
  - 3-stage以上 : 2件

Langrand-Escure J, Rivoirard R, Oriol M, et al: PLOS ONE 12:e0185536, 2017のsupplementary のexcelファイルより集計

ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

実際に多く使われているのは？：文献②

2005,2010,2014年のJCO, Ann Oncol, Cancer, CCR, Lancet Onc, NEJM, Lancetの phase 2 を調査

Table 4. Statistical design and number of patients per arm.

	2005 (n = 141)	2010 (n = 150)	2014 (n = 56)	Total (n = 347)
Multi-arm	50 (35%)	51 (34%)	28 (50%)	129 (37%)
Randomized				
Yes	31 (62%)	34 (67%)	23 (82%)	88 (25%)
No	19 (38%)	17 (33%)	5 (18%)	38 (11%)
Comparative				
Yes	20 (40%)	17 (33%)	9 (32%)	46 (36%)
No	30 (60%)	34 (67%)	19 (68%)	83 (64%)
Single arm	91 (65%)	99 (66%)	28 (50%)	218 (63%)
Statistical design				
Single stage	41 (45%)	44 (44%)	10 (36%)	95 (44%)
Simon	39 (43%)	35 (35%)	13 (46%)	87 (40%)
Fleming	2 (2%)	7 (7%)	3 (11%)	12 (6%)
Bayesian	1 (1%)	6 (6%)	0 (0%)	7 (3%)
Gehan	5 (6%)	1 (1%)	1 (3%)	7 (3%)
Flexible	0 (0%)	3 (3%)	0 (0%)	3 (1%)
Bryant & Day	2 (2%)	1 (1%)	0 (0%)	3 (1%)
Sargent	0 (0%)	2 (2%)	0 (0%)	2 (1%)
London	0 (0%)	0 (0%)	1 (4%)	1 (0%)
Triangular test	1 (1%)	0 (0%)	0 (0%)	1 (0%)

1-stageとSimonの方法が主流

Ivanova A, Paul B, Marchenko O, et al: J Biopharm Stat 26:141-9, 2016

ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

1-stage？ 2-stage ？

- TAPUR (Targeted Agent and Profiling Utilization Registry)
  - FDAで承認済みのoff-label の薬剤を評価する試験
  - 解析対象は、ある薬剤を投与された、特定の原発巣で、特定の遺伝子変異を持つ患者

- NCI-MATCH（Molecular Analysis for Therapy Choice）
  - 特定の遺伝子変異に対する薬剤の評価をする試験
  - 解析対象は、ある薬剤を投与された、~~特定の原発巣で、~~特定の遺伝子変異を持つ患者

Simon R: Ann Intern Med 165:270-8, 2016
Mangat PK, Halabi S, Bruinooge SS, et al: JCO Precision Oncology:1-14, 2018

ICRweb 臨床研究入門 www.icrweb.jp
営利目的でのご利用や無断転載はご遠慮ください

使いわけの考え方（決まったルールはない）

- 3段階以上のデザイン：ほぼ使われていない
  - 奏効有無判明までの期間の患者登録停止期間の無駄
    - ただしEDCの導入により以前よりはデメリットが小さくなりつつある
- 1-stage：有効性がわかっている薬剤の時に正当化される
  - 特定がん種に対する既承認薬の組み合わせなど
- 2-stage：1-stageで行う理由が説明しにくい場合
  - 倫理的配慮で無効中止を考慮
  - 有効中止の考慮は不要
    - 効果あるなら倫理的問題は生じない/有効性・安全性の情報の蓄積
- Simonの Optimal vs Minimax vs SWOG
  - 試験実施組織として複数試験全体でのNを抑えたいならoptimal
  - 1試験の事だけならminimax
  - 多施設で対象者数がコントロールしにくいならSWOGも選択肢

営利目的でのご利用や無断転載はご遠慮ください

補足

サンプルサイズとpositiveとなる奏効数

営利目的でのご利用や無断転載はご遠慮ください

あるコンセプト検討会での出来事

閾値＝5%、期待値＝30%、片側α＝5%、検出力＝80%で
計算したところ必要解析対象者数は14例でした。
若干の不適格例を見込んで予定登録数を17例としました

それでも良いのですが、N＝17とすると
検出力＝80%が担保できません。
18例以上にするか、N＝16にしておいた方
が良いかもしれません。

え？なんでN＝14よりも増やしているのに検出力が
落ちるんですか？何をおっしゃっているのかわかりません・・・。

ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

復習：αエラーとβエラー

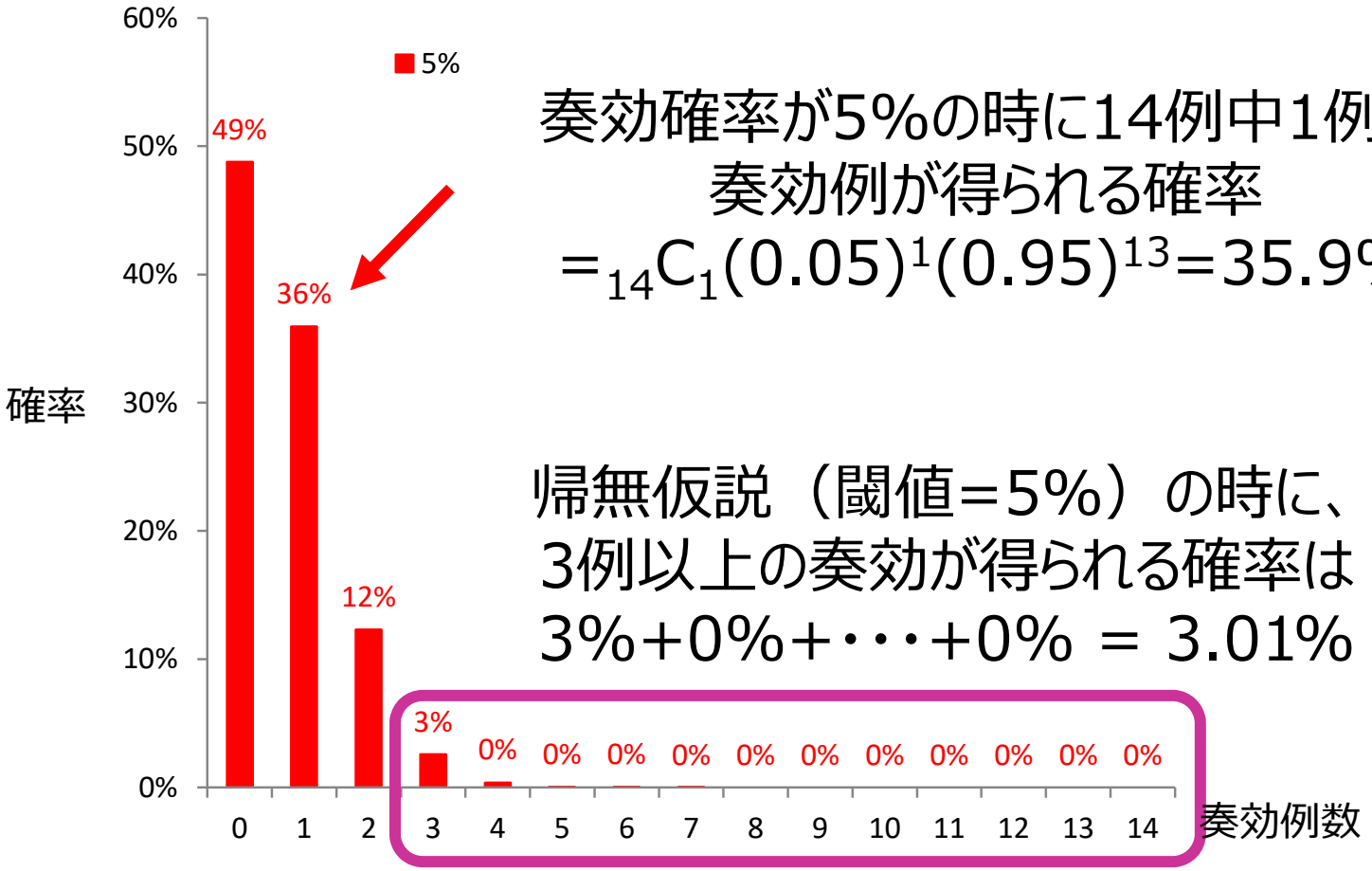
		真実	
		効果なし（閾値）	効果あり（期待値）
検定結果	有意差なし	正しい	誤り （**βエラー**）
	有意差あり	誤り （**αエラー**）	正しい （**検出力.1-β**）

- αとβは同時に小さくすることはできない
- 同時に小さくしたければNを増やす必要あり

ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

閾値5%でN=14例の時に得られる奏効例数の確率



奏効確率が5%の時に14例中1例の
奏効例が得られる確率
$={}_{14}C_1(0.05)^1(0.95)^{13}=35.9\%$

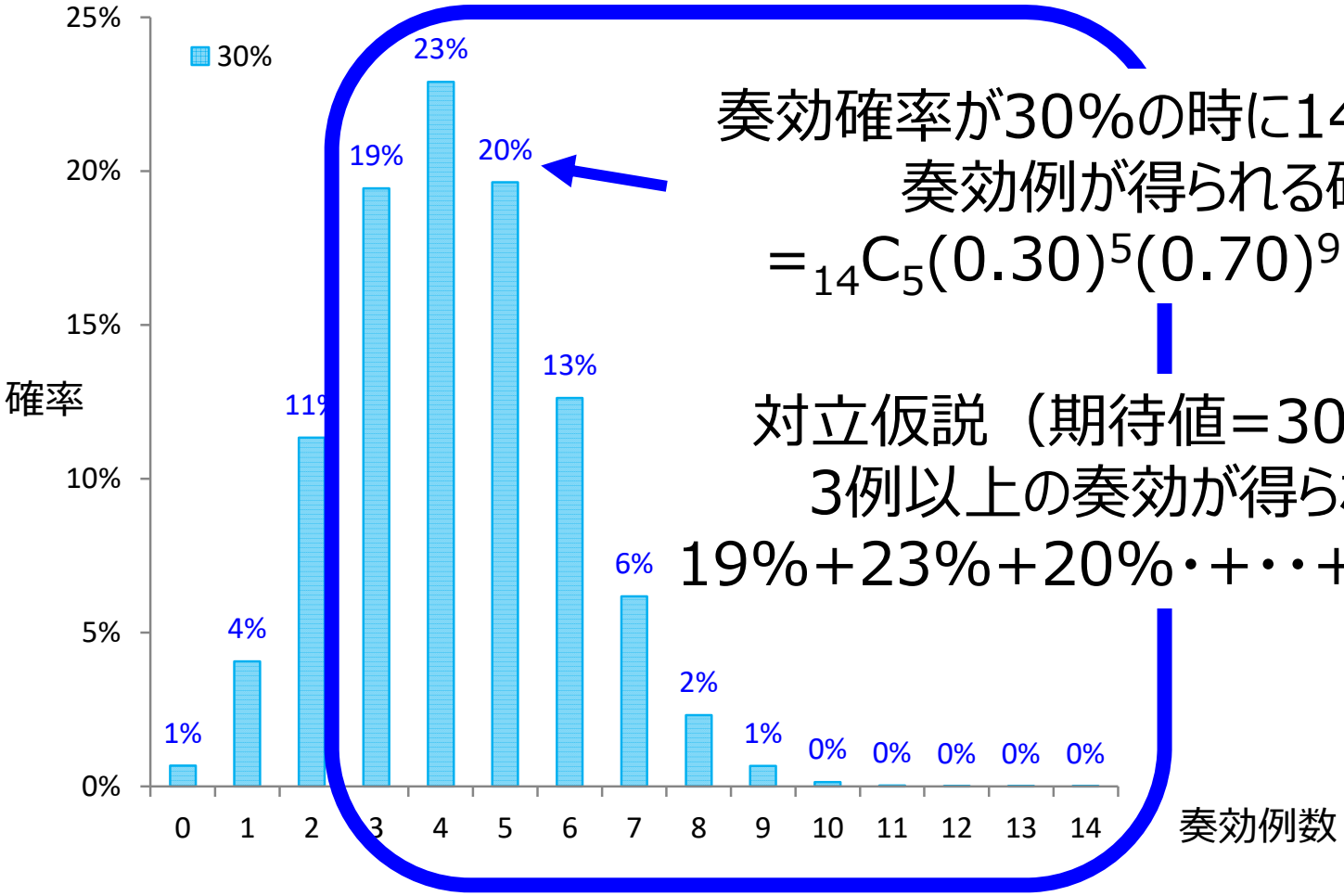
帰無仮説（閾値=5%）の時に、
3例以上の奏効が得られる確率は
3%+0%+・・・+0% = 3.01%

3例以上の奏効でpositiveと判断すればα<5%となる

ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

期待値30%でN=14例の時に得られる奏効例数の確率



奏効確率が30%の時に14例中5例の
奏効例が得られる確率
$={}_{14}C_5(0.30)^5(0.70)^9=19.6\%$

対立仮説（期待値=30%）の時に、
3例以上の奏効が得られる確率は
19%+23%+20%･+･･+0% = 83.9%

3例以上の奏効でpositiveと判断すれば検出力>80%となる

ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

同様に各Nの条件で計算する
α=5%を満たす条件での検出力

N	Positiveとなる奏効例数（奏効割合）	実際のα	実際の検出力
13	3 (23%)	2.5%	79.8%
14	3 (21%)	3.0%	83.9%
15	3 (20%)	3.6%	87.3%
16	3 (19%)	4.3%	90.1%
17	4 (24%)	0.9%	79.8%
18	4 (22%)	1.1%	83.5%
19	4 (21%)	1.3%	86.7%
20	4 (20%)	1.6%	89.2%

13例以下では検出力80%を担保できない

Nを増やすと、α<5%を満たしながらPositiveとなる奏効例数のハードルが高くなり、powerが落ちる場合がある

ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

質問：必要解析対象数の設定は？

N	Positiveとなる奏効例数（奏効割合）	実際のα	実際の検出力
13	3 (23%)	2.5%	79.8%
14	3 (21%)	3.0%	83.9%
15	3 (20%)	3.6%	87.3%
16	3 (19%)	4.3%	90.1%
17	4 (24%)	0.9%	79.8%
18	4 (22%)	1.1%	83.5%
19	4 (21%)	1.3%	86.7%
20	4 (20%)	1.6%	89.2%

① 14例

② 16例

③ 15例

④ 17例

⑤ 18例

⑥ 19例

⑦ 20例

ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

まとめ

- 閾値と期待値の意味
- P値と信頼区間の関係
- 1-stage と 2-stageデザインの使いわけ
- Simonのminimaxとoptimalの違い
- Nを増やしてもpowerが減る理由

ICRweb 臨床研究入門 www.icrweb.jp

営利目的でのご利用や無断転載はご遠慮ください

Appendix

営利目的でのご利用や無断転載はご遠慮ください

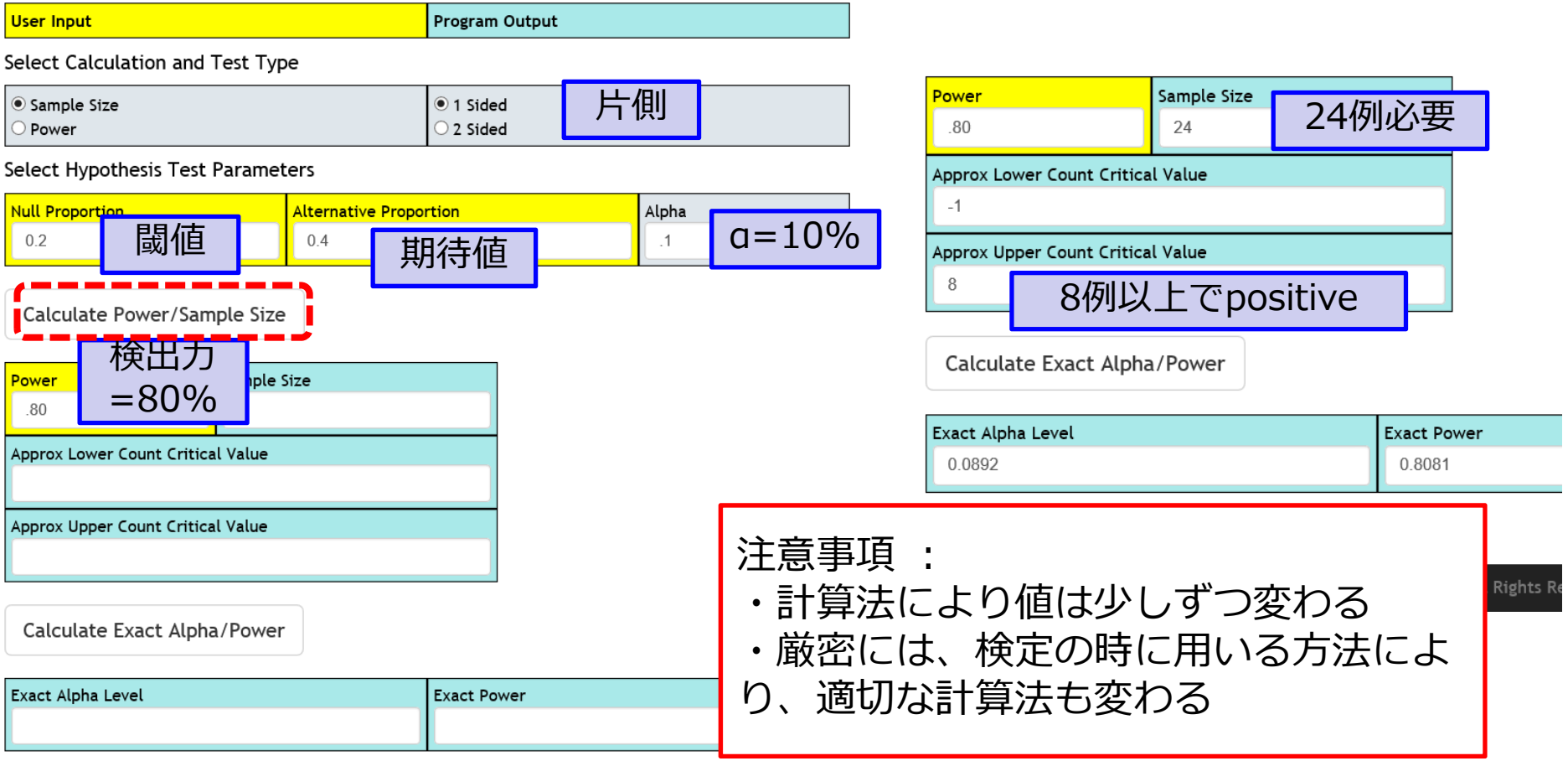
単群の割合をエンドポイントとした試験：SWOGのwebsiteの場合

- SWOGのwebsite「https://stattools.crab.org/」

営利目的でのご利用や無断転載はご遠慮ください

One Arm Binomial

- 片側α=10%,検出力=80%,閾値=20%,期待値=40%の場合

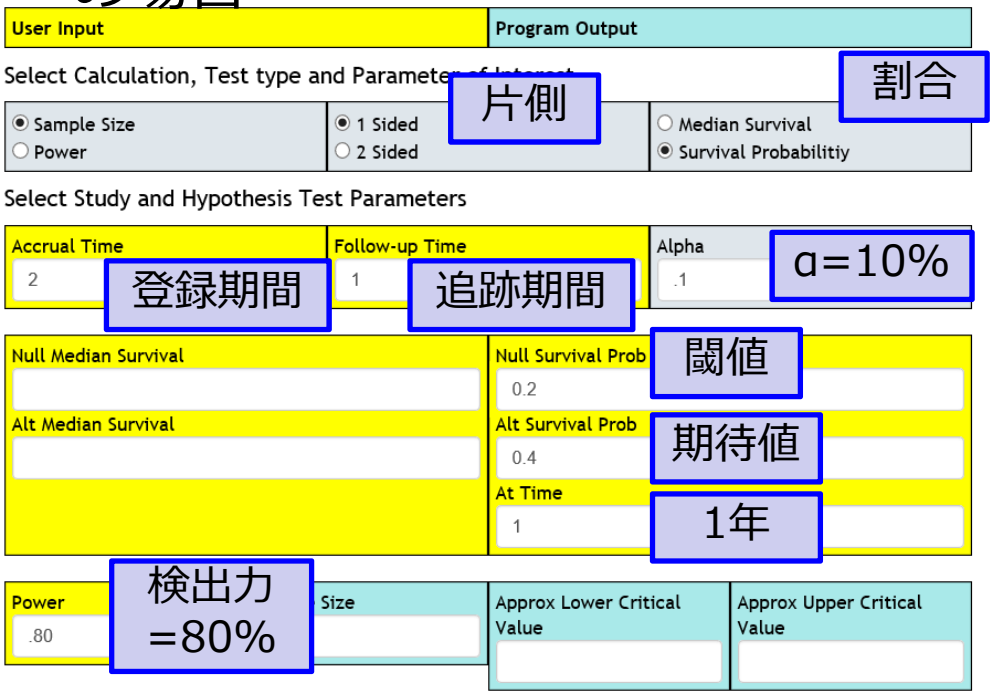


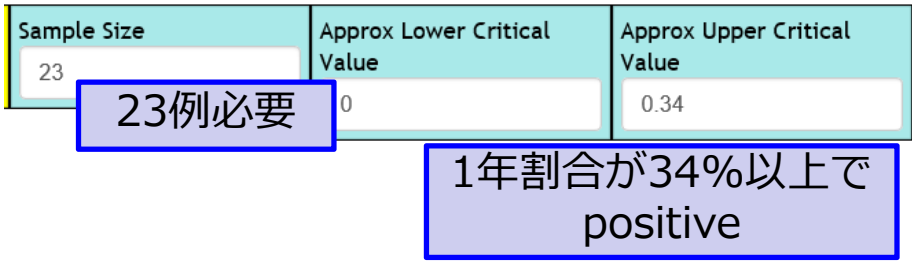
注意事項 ：
・計算法により値は少しずつ変わる
・厳密には、検定の時に用いる方法により、適切な計算法も変わる

ICRweb 臨床研究入門 www.icrweb.jp
営利目的でのご利用や無断転載はご遠慮ください

One Arm Nonparametric Survival

- 片側α=10%,検出力=80%,1年OS閾値=20%,1年OS期待値=40%の場合



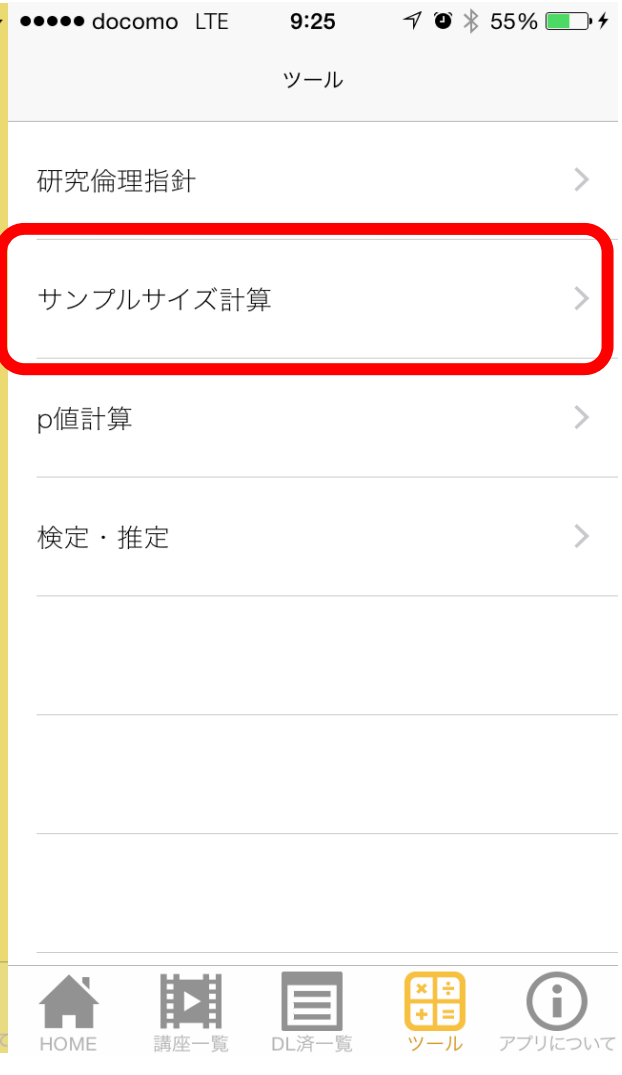


・登録期間：実際に登録できそうな期間でOK
・追跡期間：エンドポイントに設定している
期間よりも長ければOK（今回なら1年以上）

注意事項 ：
・KM法で生存割合を算出する前提

ICRweb 臨床研究入門 www.icrweb.jp
営利目的でのご利用や無断転載はご遠慮ください

単群の割合をエンドポイントとした試験：ICRwebアプリの場合



営利目的でのご利用や無断転載はご遠慮ください

割合の閾値・期待値

- 片側α=10%,検出力=80%,
  閾値=20%,期待値=40%の場合

・サンプルサイズ最小値：適当に値を設定
・サンプルサイズ最大値：実施可能な最大のNを入れておく

検出力を共に満たすパターン一覧

(閾値割合：0.2 期待値割合：0.4 片側有意水準：0.1 検出力：0.8)

項番	必要例数	この値以上ならpositive	実質α	実質power
1	24	8		
2	27	9		
3	28	9		
4	30	10	0.061087	0.823714
5	31	10	0.074600	0.856624
6	32	10	0.089832	0.884448
7	33	11	0.050837	0.831026
8	34	11	0.062027	0.861685
9	35	11		
10	36	11		
11	37	12		
12	38	12	0.062309	0.891120
13	39	12	0.074226	0.911782
14	40	12	0.087505	0.929051
15	41	13	0.052094	0.894514
16	42	13	0.062107	0.914045
17	43	13	0.073321	0.930451

30例以上であれば、確実に検出力≥80%

「必要例数」の数値の抜けがなくなる最小の値を選ぶ

補足 ：
・Nが多めになる手法を採用しているので、実施可能ならこの方法を用いればOK
・Nが多すぎたり、最小値～最大値の範囲が広いと計算時間がかかる

営利目的でのご利用や無断転載はご遠慮ください