

# Estimandの概要と がん臨床試験への適用

国立がん研究センター研究支援センター 生物統計部/  
中央病院 臨床研究支援部門 研究企画推進部 生物統計室/  
JCOGデータセンター 統計部門

水澤 純基

2025.2.13 (木) 13:15 – 14:00

東京大学大学院医学系研究科生物統計情報学講座主催 シンポジウム  
ICH E9(R1)によって日本の臨床試験の計画と解析は変わるのか-

# ICH（医薬品規制調和国際会議）

International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use（医薬品規制調和国際会議）の略称

Quality : 品質（品質に関するガイドライン）

ICH-Q1～Q14

- ICH-Q1:安定性
- ICH-Q9:品質リスクマネジメント
- ICH-Q10:品質システム
- ICH-Q7:GMP

など

Safety : 安全性（非臨床に関するガイドライン）

ICH-S1～S12

- ICH-S2:遺伝毒性試験
- ICH-S7:薬理試験
- ICH-S9:抗悪性腫瘍薬の非臨床評価
- UCH-S11:小児用医薬品開発の非臨床試験

など

Efficacy : 有効性（臨床に関するガイドライン）

ICH-E1～E20

- ICH-E3:治験報告書
- ICH-E6:GCP
- ICH-E8:臨床試験の一般指針
- **ICH-E9:臨床試験のための統計的原則**
- ICH-E17:国際共同治験

など

Multidisciplinary : 複合領域（品質・安全性・有効性の複数領域に関するガイドライン）

ICH-1～M14

- ICH-M1: 国際医薬用語集
- ICH-M14:安全性評価においてRWDを活用する薬剤疫学調査の計画・デザインに関する一般原則

など

業務別		訪問者別	製品種類別	
承認審査関連業務	安全対策業務	健康被害救済業務	レギュラトリーサイエンス・ 基準作成調査・日本薬局方	国際関係業務（ICH・IMDRF 等）

ホーム > 国際関係業務 > 国際調和活動 > ICH 医薬品規制調和国際会議 > ガイドライン > Efficacy：有効性 臨床に関するガイドライン > ICH-E9

国際関係業務

- 1998年にE9が通知
- 2019年に補遺という形でE9 (R1)が通知され、“**estimand**”という概念が導入
- 2024年にstep5になり、今後、プロトコールにも**estimand**に関連する用語が用いられる
  - 研究者主導臨床試験も治験と同様に用いるようになると思われる

### E9（ステップ5）

- [臨床試験のための統計的原則 \[118.35KB\]](#)  (1998年11月30日)
- [\(原文\) Statistical Principles for Clinical Trials \[94.59KB\]](#) 

### E9 (R1)（ステップ5）

- [臨床試験のための統計的原則 補遺 臨床試験におけるestimandと感度分析 \[415.19 KB\]](#)  (2024年6月20日)
- [\(原文\) Addendum on Estimands and Sensitivity Analysis in Clinical Trials to The Guideline on Statistical Principles for Clinical Trials.\[156.75KB\]](#) 

[ステップの説明はこちら](#)

国際関係業務

国際関係業務の概要

国際調和活動

ICMRA

ICH 医薬品規制調和国際会議

ガイドライン

Quality：品質 品質に関する  
ガイドライン

Safety：安全性 非臨床に関  
するガイドライン

# Estimandの定義

試験の目的によって提起される臨床的疑問を反映する**治療効果の詳細な説明**。比較されている異なる治療状況下において同じ患者の結果がどのようになるかを集団レベルで要約するものである。

## Estimandを構成する5つの要素

- ✓ 臨床的疑問に対応する**対象集団** **P**
- ✓ 関心のある**治療**の状況と、必要に応じて、比較を行う他の治療の状況 **I C**
- ✓ 臨床的疑問に対応するために必要な、各患者について収集する**変数**（又は**評価項目**） **O**
- ✓ 治療開始後に発現し、関心のある科学的疑問に関連する測定値の解釈や有無の影響を及ぼす事象（**中間事象**）をどのように考慮するか **?**
- ✓ **集団レベルでの要約** **O**

# estimandを構成する5つの要素

## 中間事象

増悪/再発、毒性治療中止、後治療、救済手術、緩和照射・・・  
[臨床的疑問を説明するために、これらの事象をどのように扱うか]

?

## 対象集団

[どのような被験者に対して]

P

## 治療

[どのような治療効果を]

I C

## 変数（エンドポイント）

個人レベル  
[どのような評価尺度で]

O

## 要約指標

集団レベル  
[どのような要約する]

O

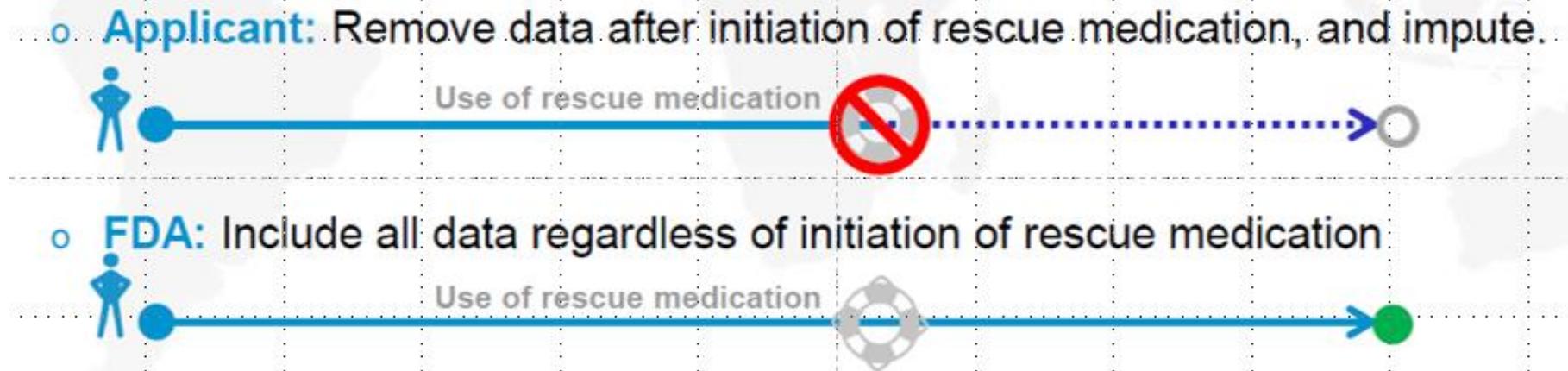
- 中間事象以外の4つはPICOと同じで今までも議論済、プロトコールや論文でも読み取れる
- 中間事象の取り扱いを明示的に考えることが今までと大きく異なる

# “治療効果”についての合意が曖昧であった例

- II型糖尿病患者を対象としたダパグリフロジンの申請
  - Primary endpoint: ベースラインから**24週後**までのHbA1cの変化 **O**
  - 対象：ランダム化された全患者 **P**
  - 治療：アクティブコントロール vs ダパグリフロジン **I C**
  - 治療中に想定されること：救済治療
    - 血糖コントロールのマーカがある閾値を超えた場合、**救済治療薬**が投与される
  - 統計解析：**救済治療薬開始後のデータは解析対象から除外**
    - 救済治療薬の摂取が治療群間の血糖値の差をなくしてしまう可能性があり、特に効果の低い治療群でより頻繁に起こるため、糖尿病の試験の解析では一般的だった
    - 当時の標準的な手法である救済治療薬前の最後の値を試験終了時の値として繰り上げて解析（LOCF）

全登録例（≡ITT解析）でPICOも明確にされているように見えるが・・・

# 中間事象の取り扱いによって興味のある治療効果が異なる



## 企業（申請者）が考えていた治療効果

- 最初にランダム化された治療法の効果に興味あり
- 24週時よりも前に救済治療を受けた患者の**24週時のデータ収集は不要**

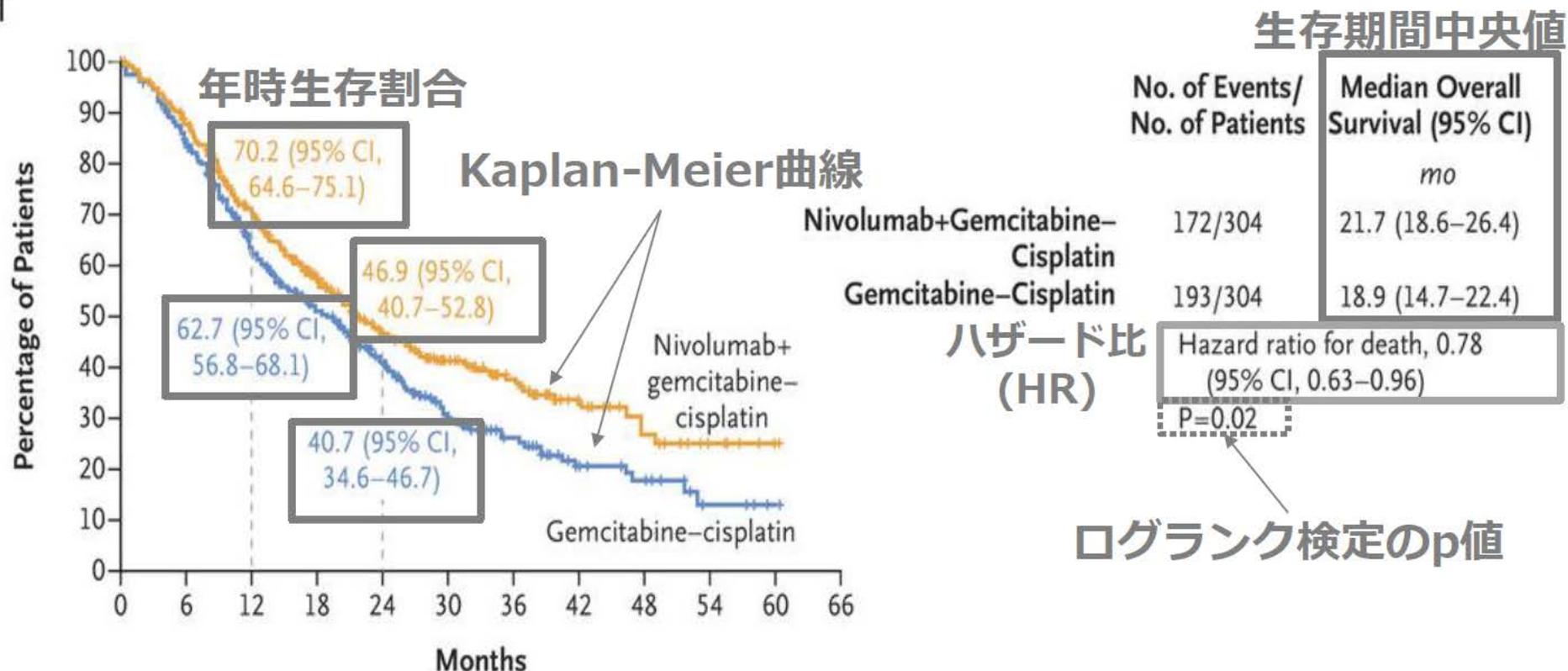
## FDA（規制当局）が考えていた治療効果

- 「ダパグリフロジン＋救済治療薬」と「コントロール＋救済治療薬」の治療方針の比較に興味あり
- 24週時よりも前に救済治療を受けた患者の**24週時のデータ収集も必要**

- どちらも**全登録例を対象**とはしているが、**中間事象**（今回の場合は救済治療）の取り扱いが違う
- 一見すると、**PICO**は両者で同じに見えるが、**中間事象**の取り扱いが異なると、興味の治療効果が異なり、認識のすり合わせができない

# がんを対象とした臨床試験の一般的な結果の示し方

## A Overall Survival



- (他領域と比較すると) **解析手法**とその**要約指標**がおおよそ決まっている
- **Kaplan-Meier法**で生存曲線を推定し**年時生存割合**や**生存期間中央値**を算出し、
- **ログランク検定**で群間比較をし、
- **Cox回帰**で治療効果の要約値として**ハザード比 (HR)**を算出する

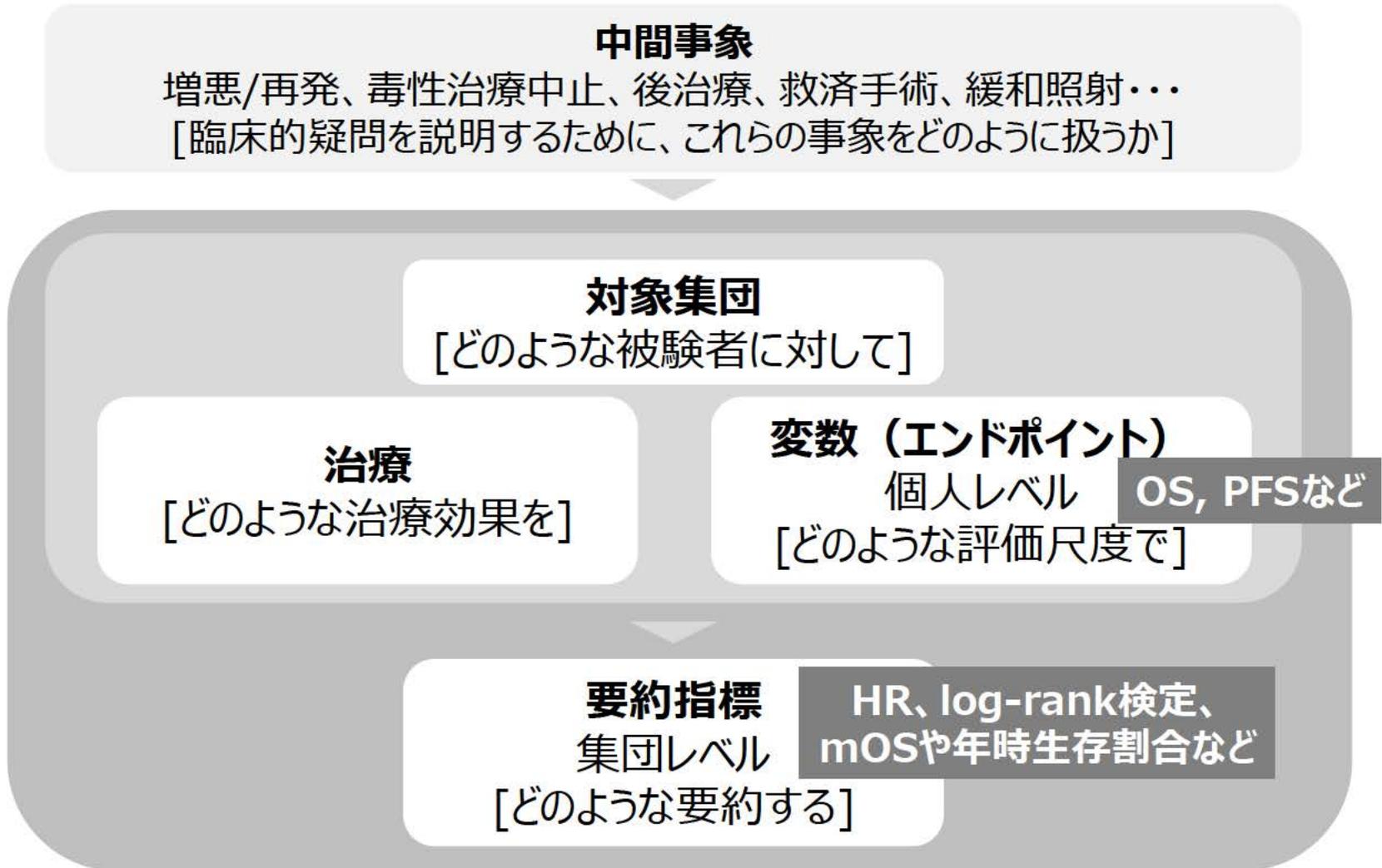
# がん領域のTime to event 型 エンドポイントの定義

- Time to event 型：ある特定の時点（起算日）から目的のイベントのうち、いずれか最も早いものが起きるまでの時間

エンドポイント	イベント（いずれか早いもの）		
<b>全生存期間 (真・ハードなエンドポイント/gold standard)</b> Overall survival (OS)	あらゆる死亡	-	-
<b>無増悪生存期間 「登録時に担がん状態」</b> Progression-free survival (PFS)	あらゆる死亡	増悪/再発	
<b>無再発生存期間 「登録時に非担がん状態」</b> Relapse-free survival (RFS)	あらゆる死亡	再発	
<b>無病生存期間 「登録時に非担がん状態」</b> Disease-free survival (DFS)	あらゆる死亡	再発	二次がん
<b>無イベント生存期間</b> Event-free survival (EFS)	血液腫瘍の場合	再発/増悪	CR/PR未達
	術前治療の場合	再発/増悪	手術不能

- よく用いられる**変数（エンドポイント）**とその定義がおおよそ決まっている
  - 個々の事象をイベント。イベントの未観測は打ち切りとして扱う

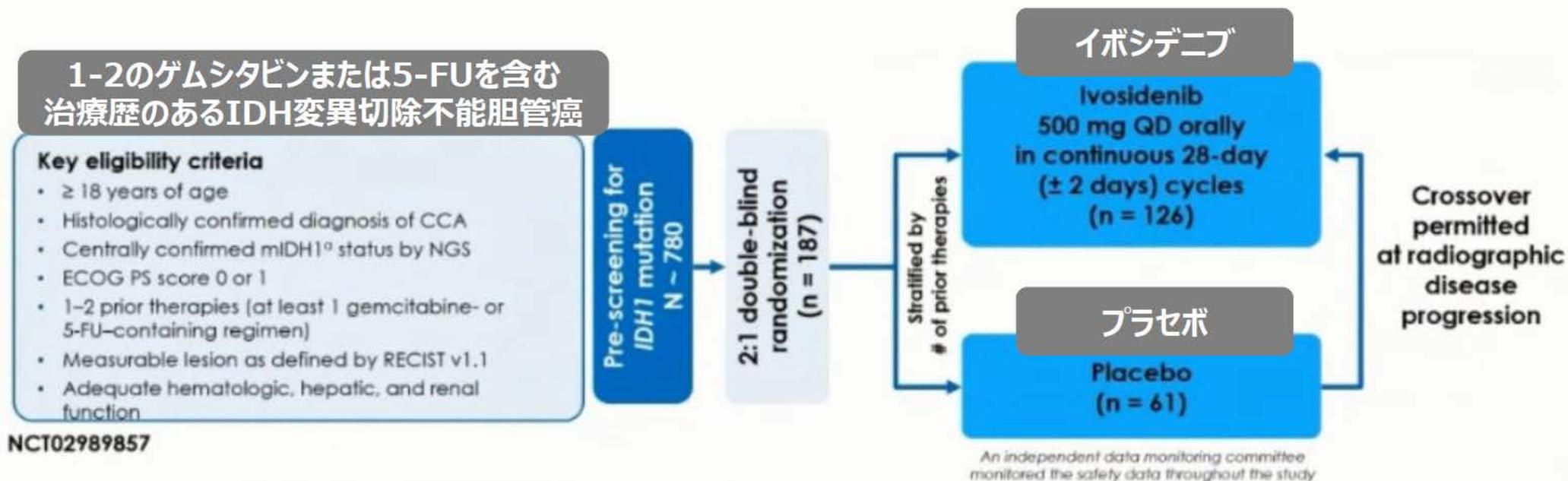
# がん領域での-estimandを構成する5つの要素-



- Estimandの一部はある意味、検討済とも言える
- がん領域ではestimandについて検討することはほとんどない？

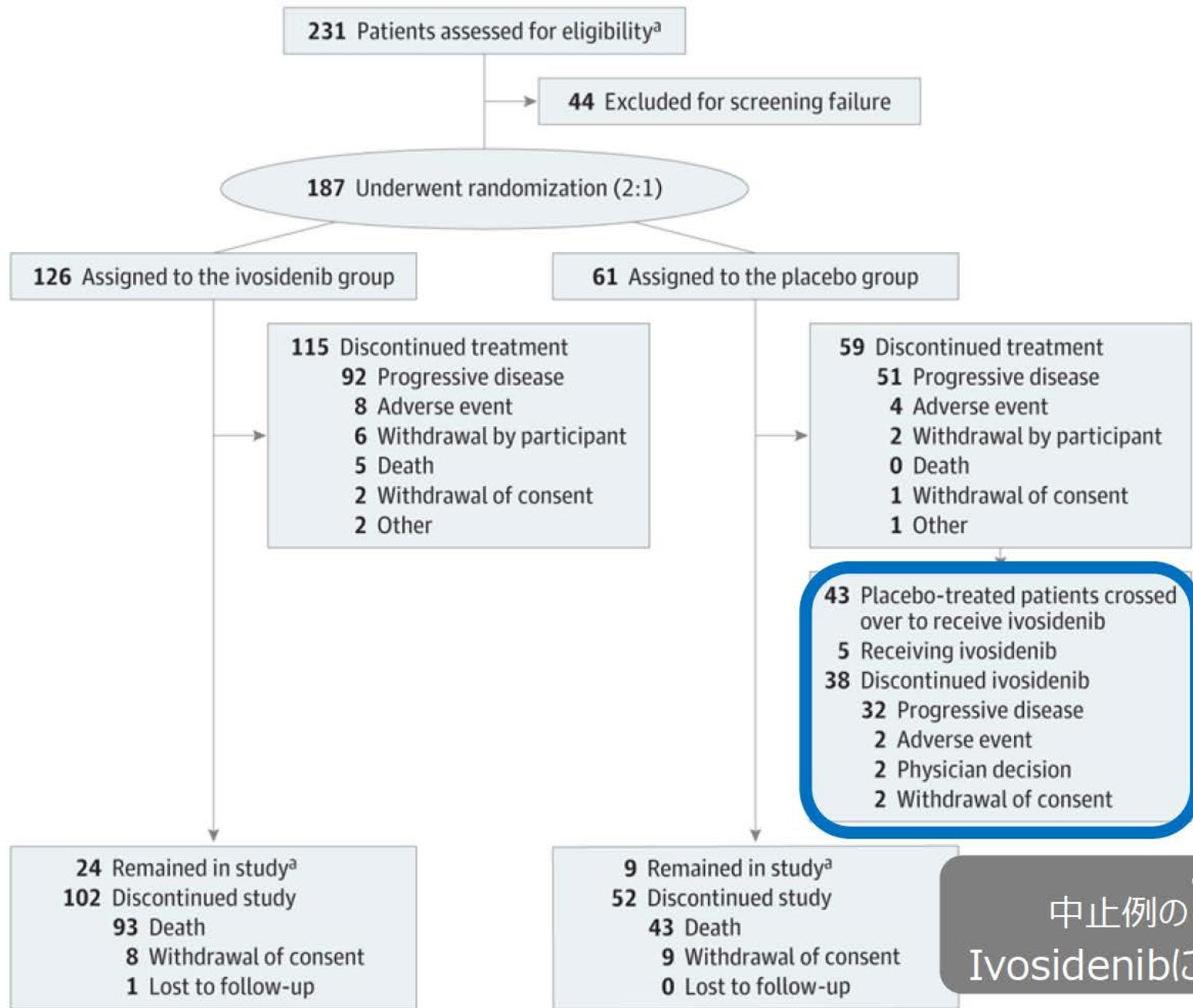
# 対照群に試験治療がクロスオーバーする場合のOS

## Phase 3 ClarIDHy study: Study design and endpoints



- **Primary endpoint:** PFS by blinded independent radiology center (IRC)
- **Key secondary endpoints:** OS; objective response rate; PFS by local review; pharmacokinetics/ pharmacodynamics; health-related quality of life (HRQOL);<sup>b</sup> safety and tolerability

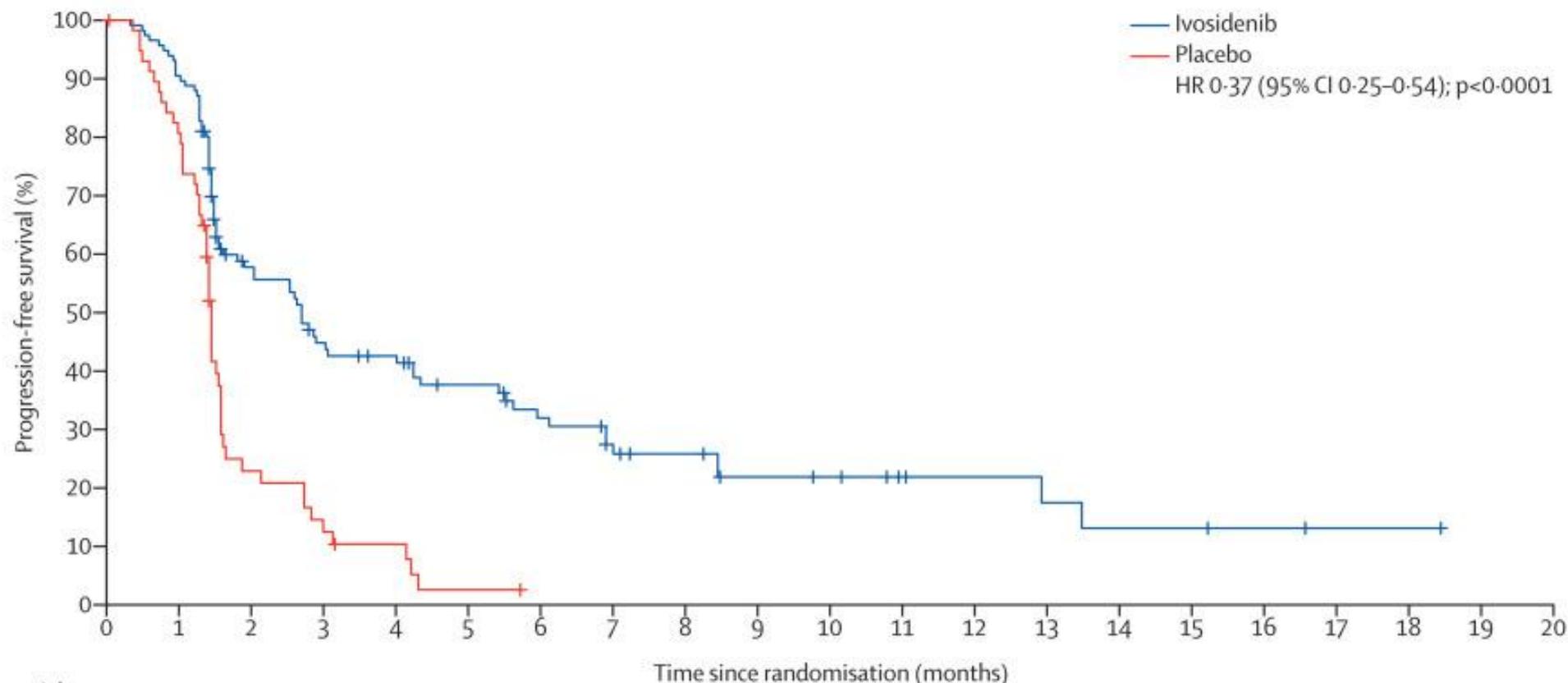
- 中央判定のPFSがprimary endpoint。OSがKey secondary endpoint
- Placebo群の患者は、増悪後、条件を満たせばIvosidenibの投与可能
- 条件を満たさない患者は有害事象中止や増悪後は、Ivosidenib以外の後治療を受ける



中止例の73%が  
Ivosidenibにクロスオーバー

# PrimaryのPFSはpositive

A

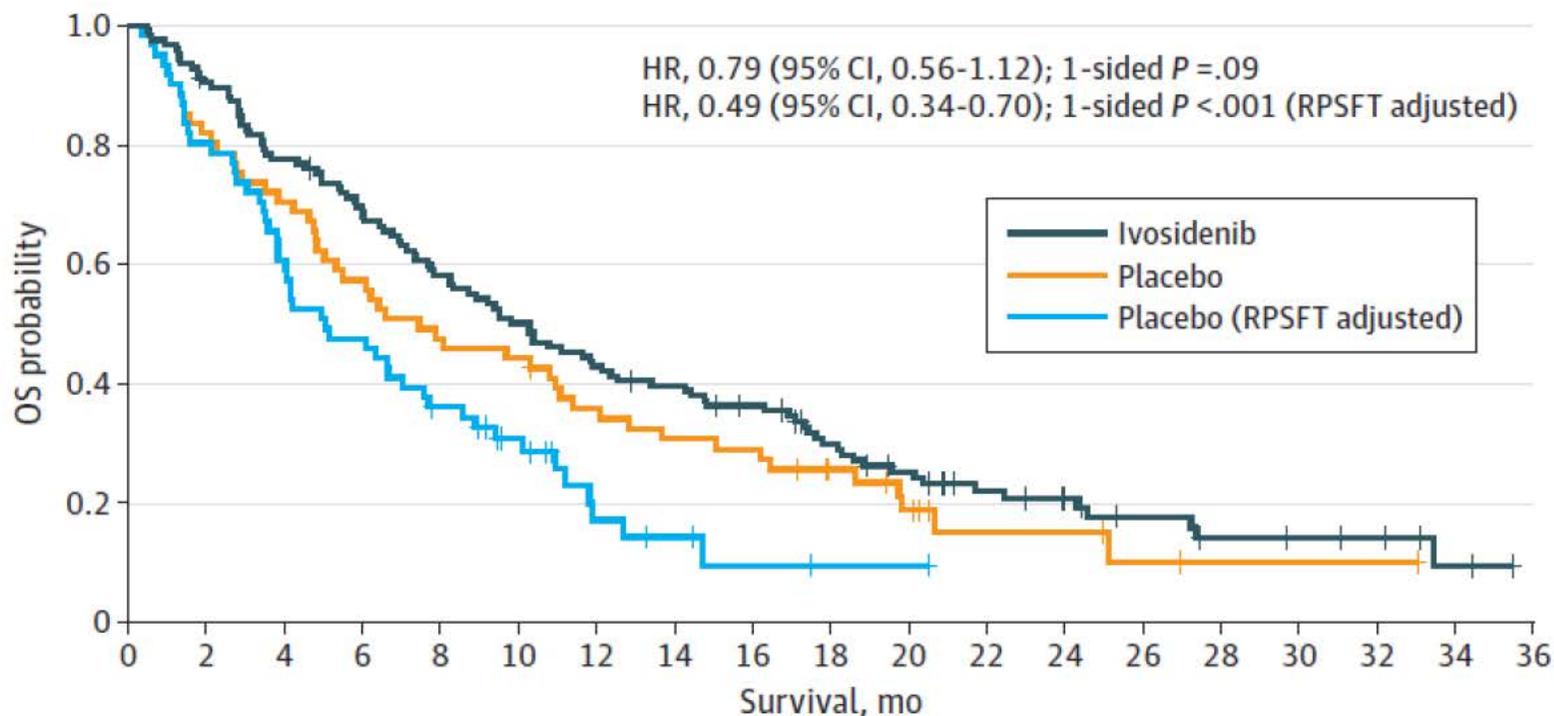


Number at risk  
(number censored)

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Ivosidenib	124 (0)	105 (8)	54 (24)	40 (26)	36 (28)	28 (32)	22 (34)	16 (36)	14 (38)	10 (40)	9 (41)	6 (44)	5 (45)	4 (45)	3 (45)	3 (45)	2 (46)	1 (47)	1 (47)	0 (48)	..
Placebo	61 (0)	46 (4)	11 (9)	6 (9)	4 (10)	1 (10)	0 (11)	..	..	..	..	..	..	..	..	..	..	..	..	..	..

Ivosidenib in IDH1-mutant, chemotherapy-refractory cholangiocarcinoma (ClarIDHy): a multicentre, randomised, double-blind, placebo-controlled, phase 3 study. *Lancet Oncol.* 2020 Jun;21(6):796-807. doi: 10.1016/S1470-2045(20)30157-1. Epub 2020 May 13. Erratum in: *Lancet Oncol.* 2020 Oct;21(10):e462. PMID: 32416072; PMCID: PMC7523268.

**A** Overall survival



No. at risk  
 Ivosidenib  
 Placebo

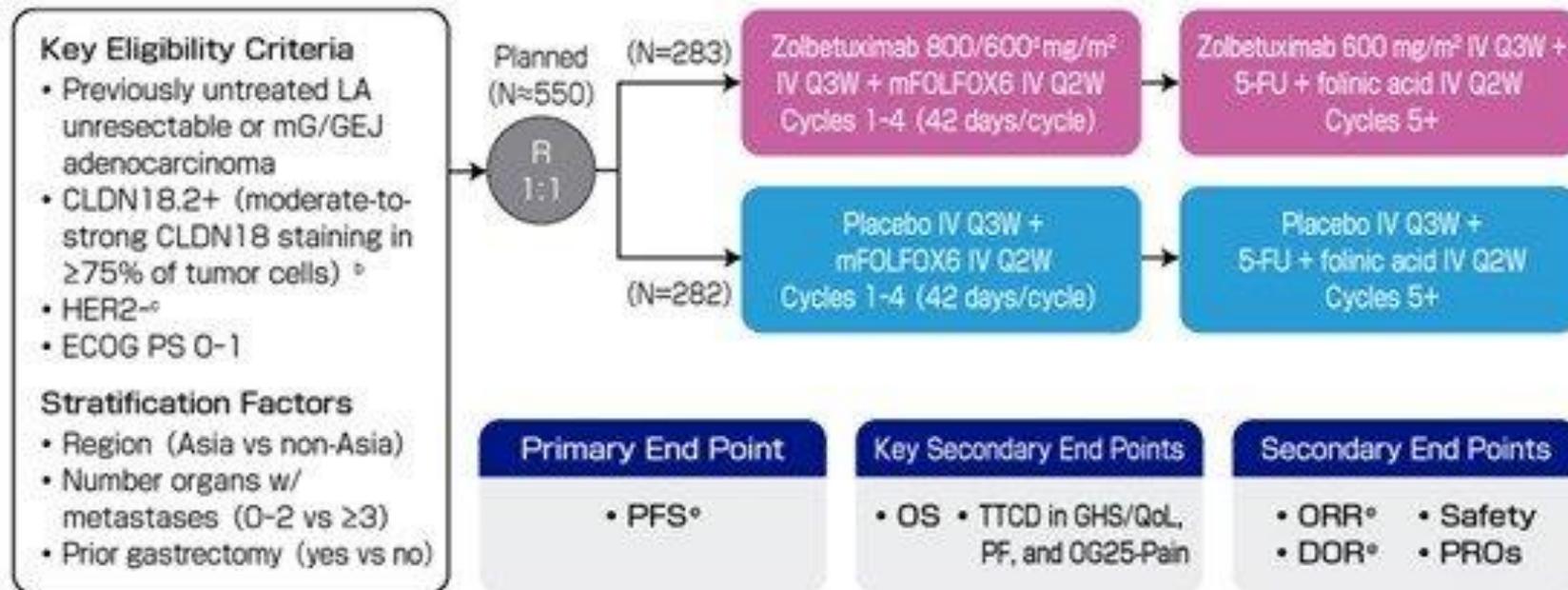
	126	113	97	85	72	62	53	48	42	32	25	18	14	10	7	6	5	2	
Ivosidenib																			
Placebo	61	50	43	35	29	27	21	18	17	12	8	4	4	2	1	1	1		

- クロスオーバーを考慮しないOSの解析はnegative
  - Ivosidenibがプラセボ群にクロスオーバーしたことで治療効果が薄まってしまった？
  - これでIvosidenibにOSの効果がない、というのは納得できる？
- 大雑把に言えば、本当に見たい治療効果は「クロスオーバーしなかった場合」なのか「クロスオーバーも含めた」はクロスオーバーという**中間事象**の取り扱いにより変わる

# PFSに関する中間事象の取り扱い

## 図1 ■ Study Design: SPOTLIGHT

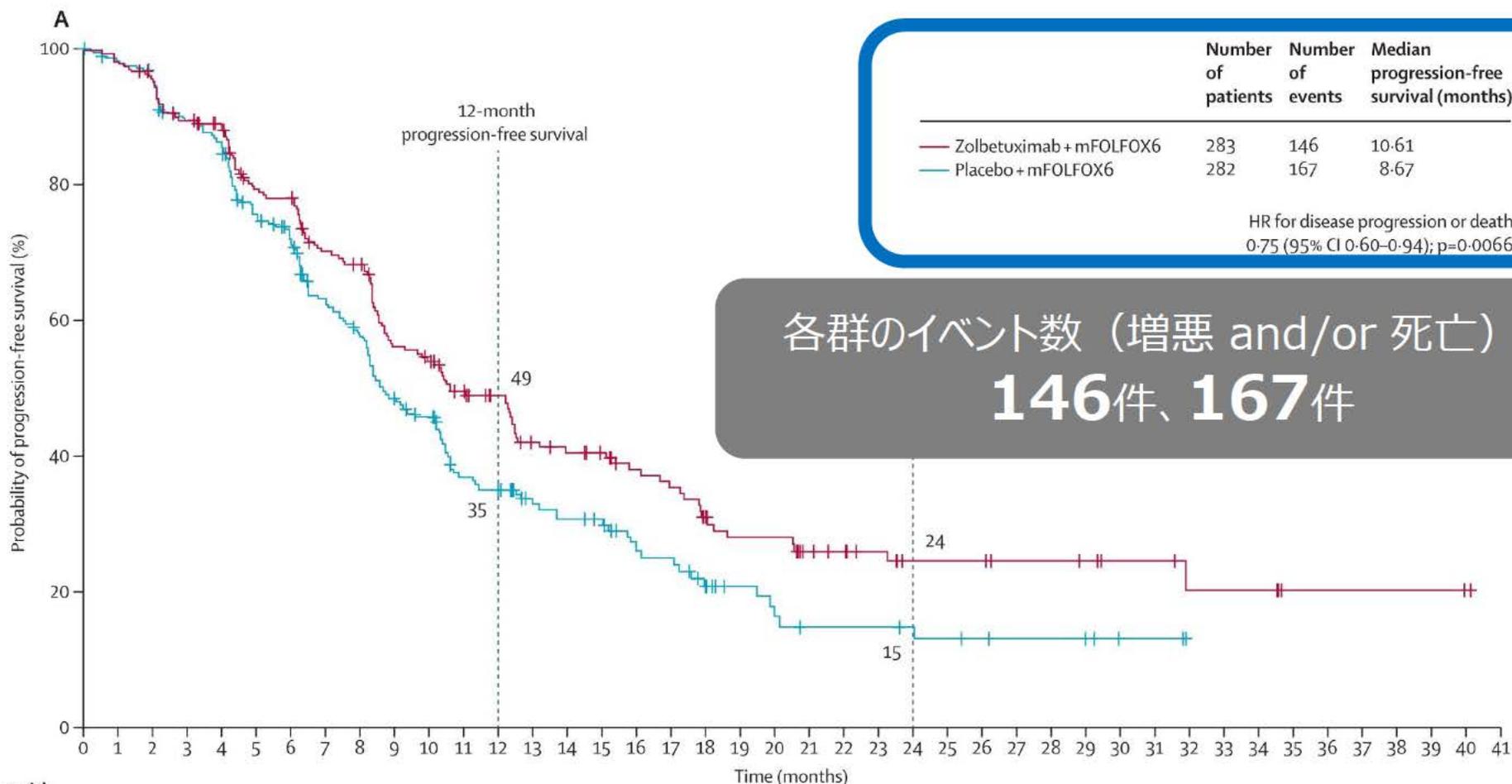
Global <sup>a</sup>, randomized, double-blinded, placebo-controlled, phase 3 trial



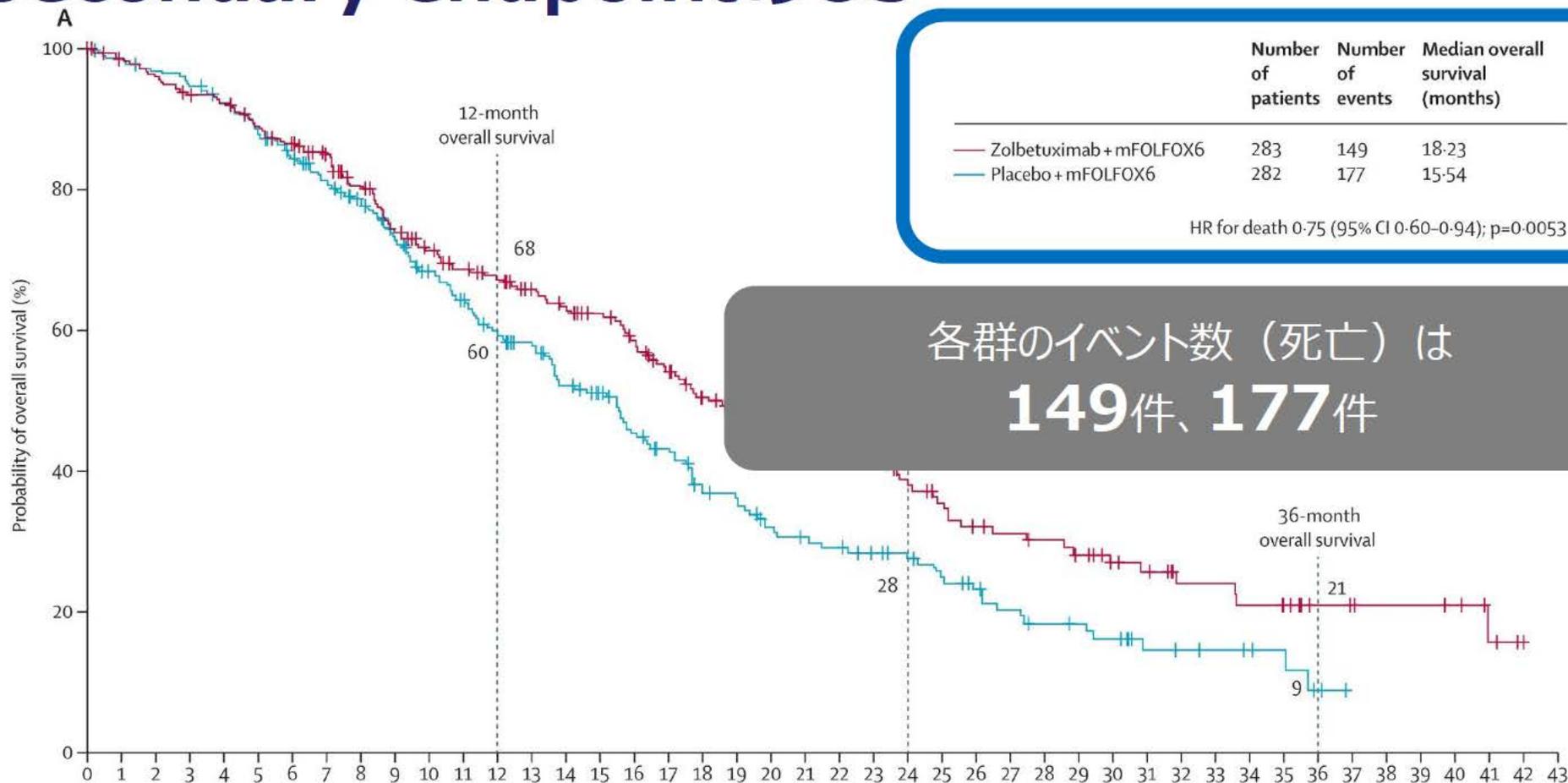
<sup>a</sup> Study was conducted at 215 sites in 20 countries across Australia, Asia, Europe, N. America, and S. America; <sup>b</sup> By central IHC using the analytically validated VENTANA CLDN18 (43-14A) RxDx Assay; <sup>c</sup> By central or local HER2 testing; <sup>d</sup> 800 mg/m<sup>2</sup> at cycle 1 day 1 followed by 600 mg/m<sup>2</sup> on cycle 1 day 22 and days 1 and 22 of subsequent cycles; <sup>e</sup> Per RECIST v1.1 by independent review committee.

(Kohei Shitara et al. ASCO GI 2023, LBA292)

# Primary endpointのPFS by 中央判定



# Secondary endpointのOS



	Number of patients	Number of events	Median overall survival (months)
Zolbetuximab + mFOLFOX6	283	149	18.23
Placebo + mFOLFOX6	282	177	15.54

HR for death 0.75 (95% CI 0.60-0.94); p=0.0053

各群のイベント数（死亡）は  
**149件、177件**

## Number at risk

Zolbetuximab+ mFOLFOX6 283 270 264 255 251 241 233 217 196 178 164 152 146 135 125 117 107 93 83 75 70 67 62 58 49 42 34 32 30 27 23 20 15 15 13 13 9 8 7 7 6 4 1 0

- OS（イベント：死亡）の数よりもPFS（イベント：増悪、死亡）の方が多
- 通常は増悪をイベントにする分だけPFSの方がイベント数は多いはず

# 無増悪生存期間(PFS)の中間事象の扱いの違い

- **共通の定義（論文にはここまでしか記載されない）**
  - 登録日を起算日とし、増悪と判断された日またはあらゆる原因による死亡日のうち早い方までの期間
  - 増悪と判断されていない生存例では臨床的に増悪がないことが確認された最終日（最終無増悪生存確認日）をもって打ち切り

## 中間事象

## SPOTLIGHT試験

## 研究者主導臨床試験(JCOGの場合)

増悪前の  
後治療開始

### 最終無増悪確認日で打ち切り

- プロトコル治療中止しているので、その後の効果判定は規定できないため薬効評価ができない

### イベントにも打ち切りにもしない

- 死亡のイベントは観察できるし、増悪判定も少なくとも日常臨床レベルで画像評価してもらえばOK

2回以上  
効果判定  
未実施後の増悪

### 最終無増悪確認日で打ち切り

- 発見時よりもっと早く増悪していた可能性あり
- ただ、それがいつかはわからないので打ち切り？

### 増悪確認日でイベント

- 奏効割合判定に必要な期間を超えると効果判定有無を確認していないためわからない

- **中間事象（増悪前の後治療開始など）の扱いによって、単純に「PFS」といっても見たい治療効果が変わる**
- **イベント有無にも影響するので、サンプルサイズ計算にも考慮が必要かも**

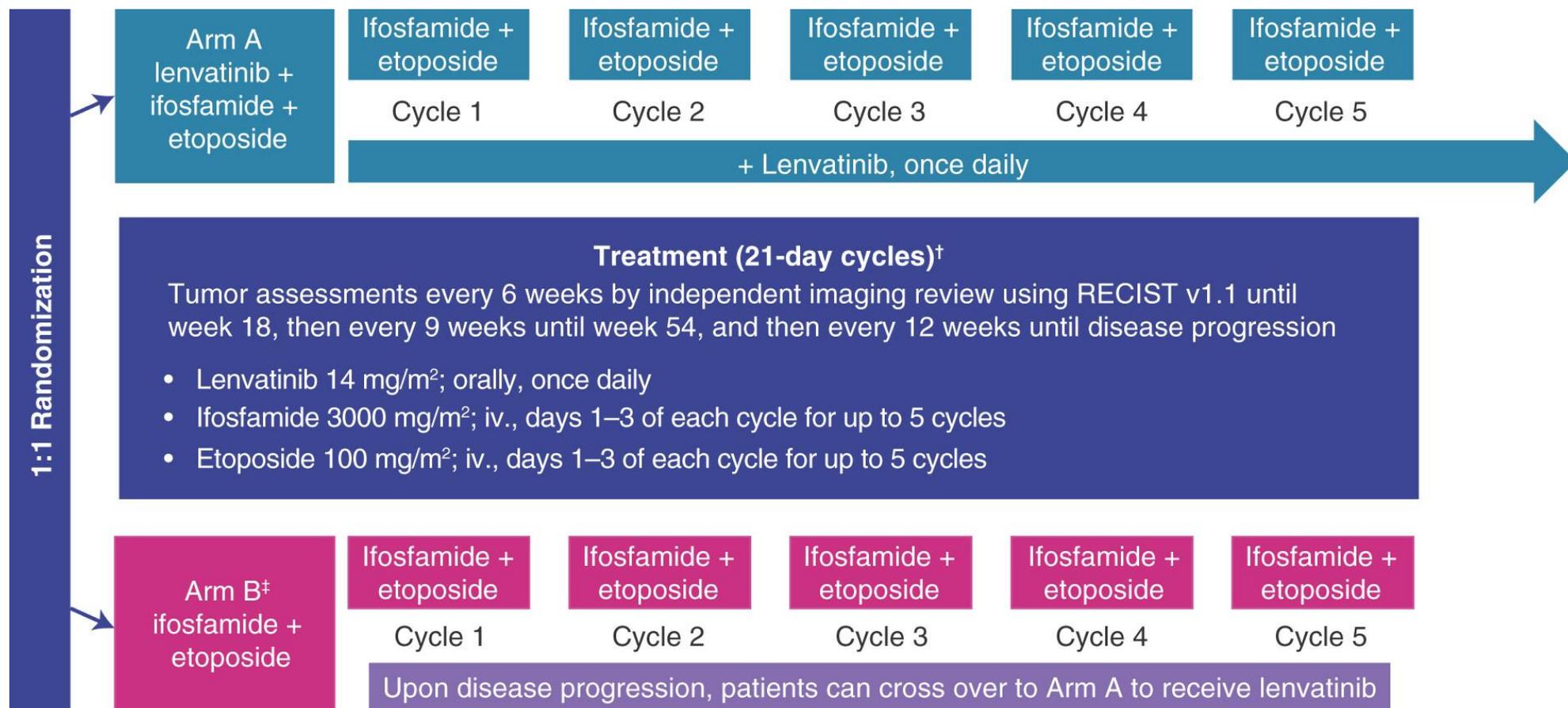
JCOGプロトコルマニュアル <https://jcog.jp/doctor/tool/manual/>

Shitara K, et al. Lancet. 2023 May 20;401(10389):1655-1668. doi: 10.1016/S0140-6736(23)00620-7. Epub 2023 Apr 15. Erratum in: Lancet. 2023 Jul 22;402(10398):290. doi: 10.1016/S0140-6736(23)01481-2. Erratum in: Lancet. 2024 Jan 6;403(10421):30. doi: 10.1016/S0140-6736(23)02762-9. PMID: 37068504.

United States Food and Drug Administration. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics: Guidance for Industry. United States Food and Drug Administration; 2007. United States Food and Drug Administration. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics: Guidance for Industry. United States Food and Drug Administration; 2018.

2回以上効果判定未実施後の増悪、は中間事象ではないと考えることもできるが、定義が異なるという観点で含めている。

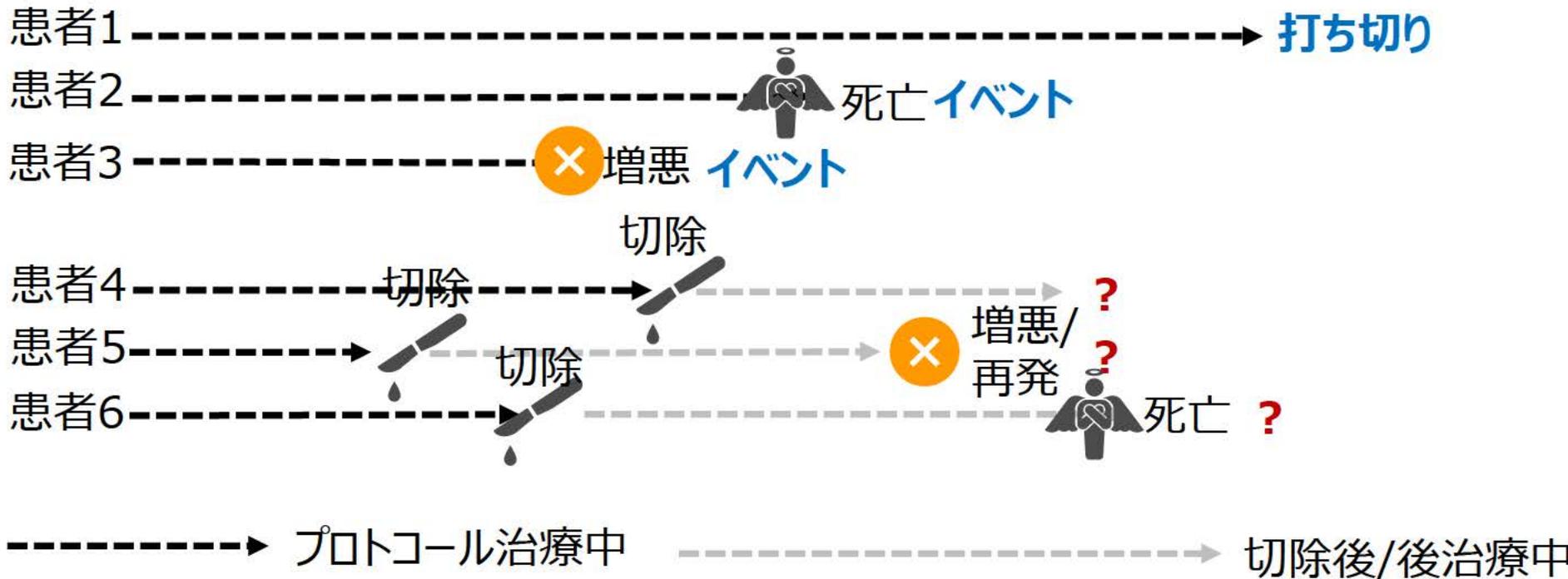
# 再発または難治性骨肉腫に対するRCT



- 試験治療はLenvatinibの上乗せ、N=80例、非盲検、承認申請が目的
- 中央判定のPFSがprimary endpoint
- 治療が効いて切除可能、あるいは切除すべきと臨床家が判断した場合は外科的切除がなされる
  - 一般論として切除すれば予後良好と考えられ、全体の10-20%くらいが切除される見込み

# 具体的に考えてみる

増悪/再発/死亡  
するまでの期間(PFS)



- プロトコール治療中の切除は治療効果の解釈に影響する
  - しかも一般的なPFSの定義ではどう扱うべきかコンセンサスはない
- この場合、切除は“**中間事象**”となり、どのように扱うかを定める必要あり
- 試験の目的/どんな治療効果を見たいのか？によって扱いが変わる
- どのように扱うかによって、データの取り方が変わる

# 中間事象の扱い方として少なくとももあり得る5つの方針

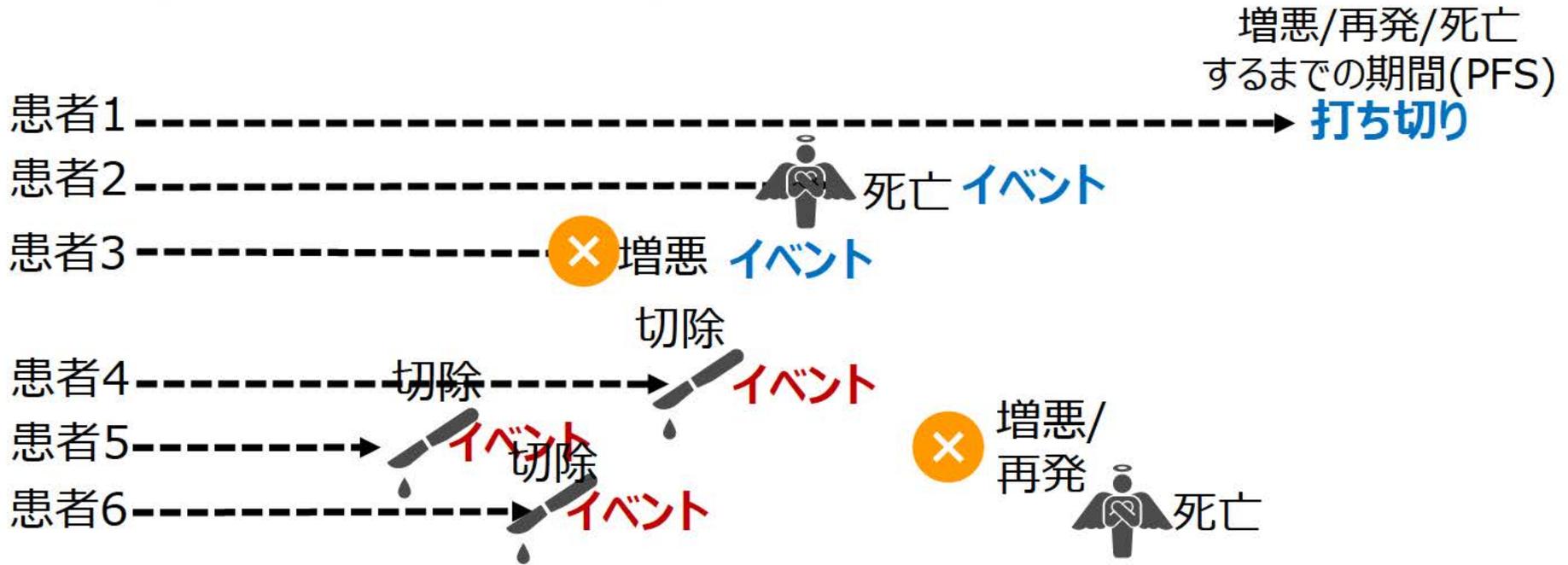
中間事象対応方針	説明
治療方針戦略 Treatment policy	<ul style="list-style-type: none"><li>中間事象発現有無によらず、関心のある変数の値を利用</li><li>中間事象発生後のイベントも興味の対象であり、評価継続が可能</li></ul>
複合戦略 Composite	<ul style="list-style-type: none"><li>中間事象発現はイベント</li><li>中間事象発現後の評価継続が実現可能でない場合の選択肢</li></ul>
仮想戦略 Hypothetical	<ul style="list-style-type: none"><li>中間事象が発現しなかった仮想的な状況での変数の値を想定</li><li>中間事象発現後の評価継続が実現可能でない場合の選択肢</li></ul>
治療下戦略 While on treatment	<ul style="list-style-type: none"><li>中間事象発現前までの治療に対するアウトカムのみを利用</li><li>中間事象発現のタイミングは無関係の時の選択肢 (例：単位時間あたりのイベント発症率など)</li></ul>
主要層戦略 Principal stratum	<ul style="list-style-type: none"><li>ランダム化前の因子で規定される中間事象が発現しない集団に関心</li><li>検証的試験で実際に使える場面はほぼ想定できない</li></ul>

# 治療方針ストラテジー：中間事象発現は無視



- 興味のある治療効果は「プロトコール治療 + 切除」であり、切除の効果も含めた治療効果に興味がある場合の対応
- **後治療後のデータ収集も必須**のため、プロトコールにそのように規定する必要がある、EDC上のデータ収集もそれに対応した設計が必要になる
- とはいっても、実臨床で非現実的であれば、規定していても実質的に意味がないので、事前にモニター、医師、CRCなどに確認が必要

# 複合ストラテジー：中間事象発現はイベント

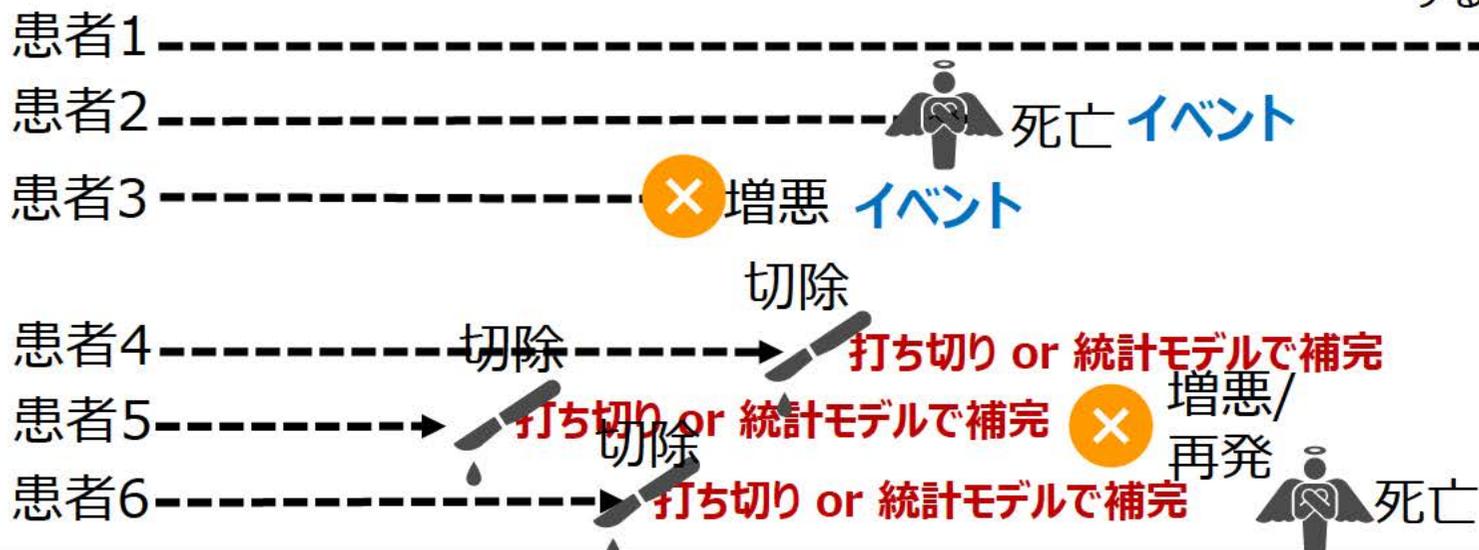


- 切除は「好ましくない事象が生じた」と考えてエンドポイントに組み込む
- 切除後のデータ収集（評価継続）が不可能な場合の選択肢として用いられる
- ただし、これが臨床的に妥当かどうかは疾患やエンドポイントの種類によるため、臨床的観点から医師に受け入れられない場合もある
  - 今回の例の場合、切除は好ましいことであるためイベントにすることは考えにくい

# 仮想ストラテジー：打ち切り or 中間事象が生じなかった

## 場合のアウトカムを統計モデルで補完

増悪/再発/死亡  
するまでの期間(PFS)



- 中間事象がなかった場合に得られる仮想的な状況での「プロトコル治療」の治療効果に関心
  - 経験上、臨床家がこれを評価したいという状況も頻繁にあり
- 打ちきりにするなら中間事象発生後のデータ収集は不要
- 単純には中間事象発生時で「打ち切り」という扱いになるが、欠点も多い
  - FDAガイダンスでは後治療開始時で打ち切り
  - 打ち切りが妥当な条件は、中間事象発生例が他の追跡継続されている患者と増悪のリスクが同じとみなせること
- 統計モデルを利用して治療効果推定をすることも選択肢だが、その統計モデルが妥当であること  
条件（測定されていない交絡がない）が強いため、批判の対象となる

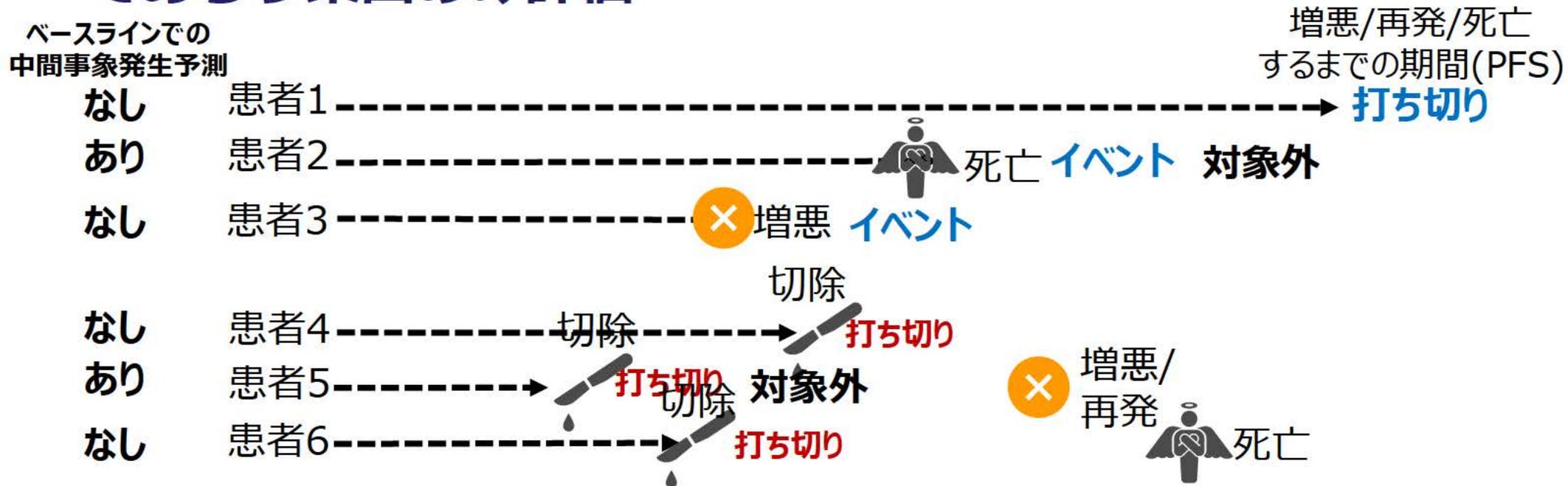
# 治療下ストラテジー：中間事象発現前までの効果を評価



- 中間事象が発生するまでの期間における治療効果に興味あり
  - がん領域だと奏効割合はこのストラテジーに対応していると考えられる
  - 緩和ケア（死亡までの期間）、毒性中止（治療中止後の追跡不可の場合）、外科切除/移植術（治癒の可能性あり、施術後追跡不可の場合）
- 中間事象発生時で「打ち切り」となり、中間事象発生後のデータ収集は不要
- 中間事象発生のリスクが高いほど目的の事象の発症率が低く見えてしまう
- 中間事象が治療と関連するかどうかで原因特異的ハザードか競合リスクを考慮した解析を用いる

# 主要層ストラテジー：関心のある中間事象が発生しない

## であろう集団のみ評価



- 割り当てられた治療によらず、関心のある中間事象が発生しないであろう集団の治療効果に興味あり
  - 今回の例で言えば、切除されない患者のサブグループでの治療効果を評価
- 一般的にはベースラインの背景因子から中間事象を発症する患者を予測し、除外
  - 実際に中間事象が生じた患者を除外するのではないことに注意
  - 予測では中間事象発生しないものの、実際には中間事象発症した場合は打ち切り
- 中間事象発生しても「打ち切り」となり、中間事象発生後のデータ収集は不要
- 統計モデルによる予測の精度によって結果の信頼性が変わるため、普通は用いられない

# 中間事象を扱うストラテジーの考え方の例

試験の目的と興味のある  
推定対象を決める

興味のある  
治療レジメン

最初に割り付けられた治療

中間事象の  
扱い

交絡として扱う

中間事象の  
扱い

各患者の変数に落とし込む

仮想ストラテジー

- 治療方針ストラテジー以外はestimandを構成する要素のいずれかに落とし込むことに相当
- 基本的に目的と興味ある推定対象が決まれば、自動的に決まるはず

治療レジメン全体  
(最初に割り付けられた治療含む)

治療方針ストラテジー

アウトカムとして扱う  
(中間事象でなくしてしまう)

複合ストラテジー

治療に落とし込む

治療下ストラテジー

集団に落とし込む

主要層ストラテジー

# まとめ

- **Estimand : 推定しようとする治療効果の詳細**
  - 対象、治療、変数、集団レベルでの要約、中間事象の5つの要素
  - 中間事象は少なくとも5つのストラテジーあり
- **がん領域はestimandの一部はコンセンサスが取れていたと考えられているかもしれないが、試験の目的に照らし合わせて考える必要あり**
  - 例えばPFSに関連する中間事象の定義を具体的に考えることでデータの収集や、試験の質をどこまで担保すべきかという議論/認識合わせにつながる