

Estimand

【第6回データサイエンスラウンドテーブル会議】

2019年3月11日

浅野淳一(PMDA)

棚瀬貴紀(大鵬薬品)

谷口隆司(ヤクルト)

手島梨恵(ノバルティスファーマ)

本テーマのねらい

- 現在、ICH E9(R1) step 2ガイドライン案とトレーニング資料が公開中
- トレーニング資料にある”thinking process”を理解し、そのprocessに沿ったestimandの構成を議論する
- パネルディスカッションを通して、産官学の立場の違いから現在関心のある論点について議論を行う

タイムテーブル

時間	内容
13:45～14:05(20分)	全体説明
14:05～15:35(90分)	グループディスカッション
15:35～15:50(15分)	休憩
15:50～16:25(35分)	グループディスカッションの結果発表
16:25～17:15(50分)	パネルディスカッション 司会:土屋悟(大日本住友製薬)、菅波秀規(興和) パネリスト:田中司朗(京都大) 田栗正隆(横浜市大) 篠崎智大(東京大) 浅野淳一(PMDA) 松岡伸篤(ファイザー) 中西豊支(興和) 棚瀬貴紀(大鵬薬品)

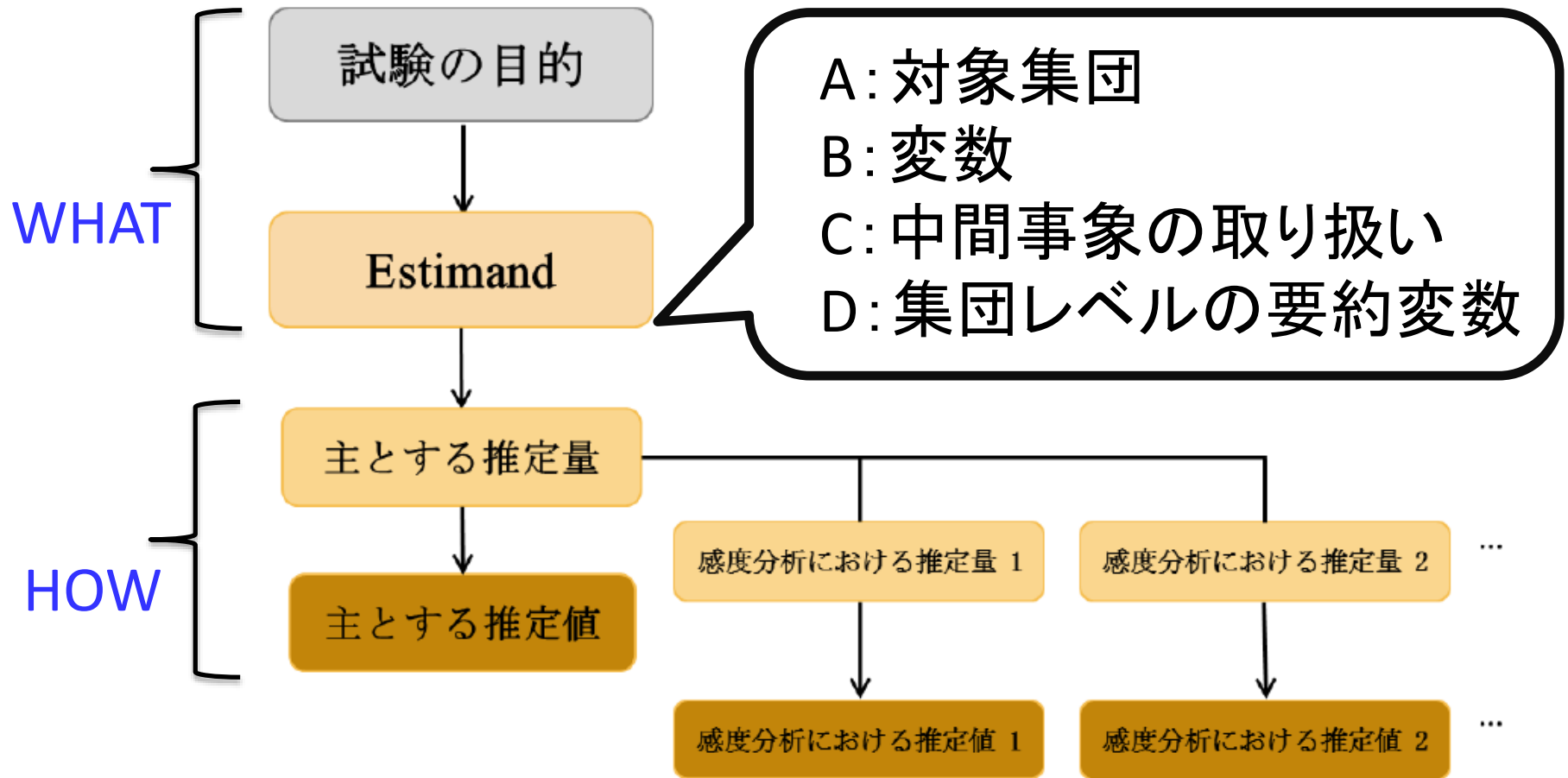
ICH E9 (R1)文書 & トレーニングマテリアル

ICH E9(R1)文書

- 2017年9月: Step 2文書公開
 - Step 3にてパブリックコメント収集
- 2018年8月: トレーニング資料公開
 - Step 2文書に従った内容. Step 4で改訂*
 - Module 1(要約)
 - Module 2(Step 2補遺文書の内容)
 - Module 3(例)
- 2019年: 補遺文書の最終化

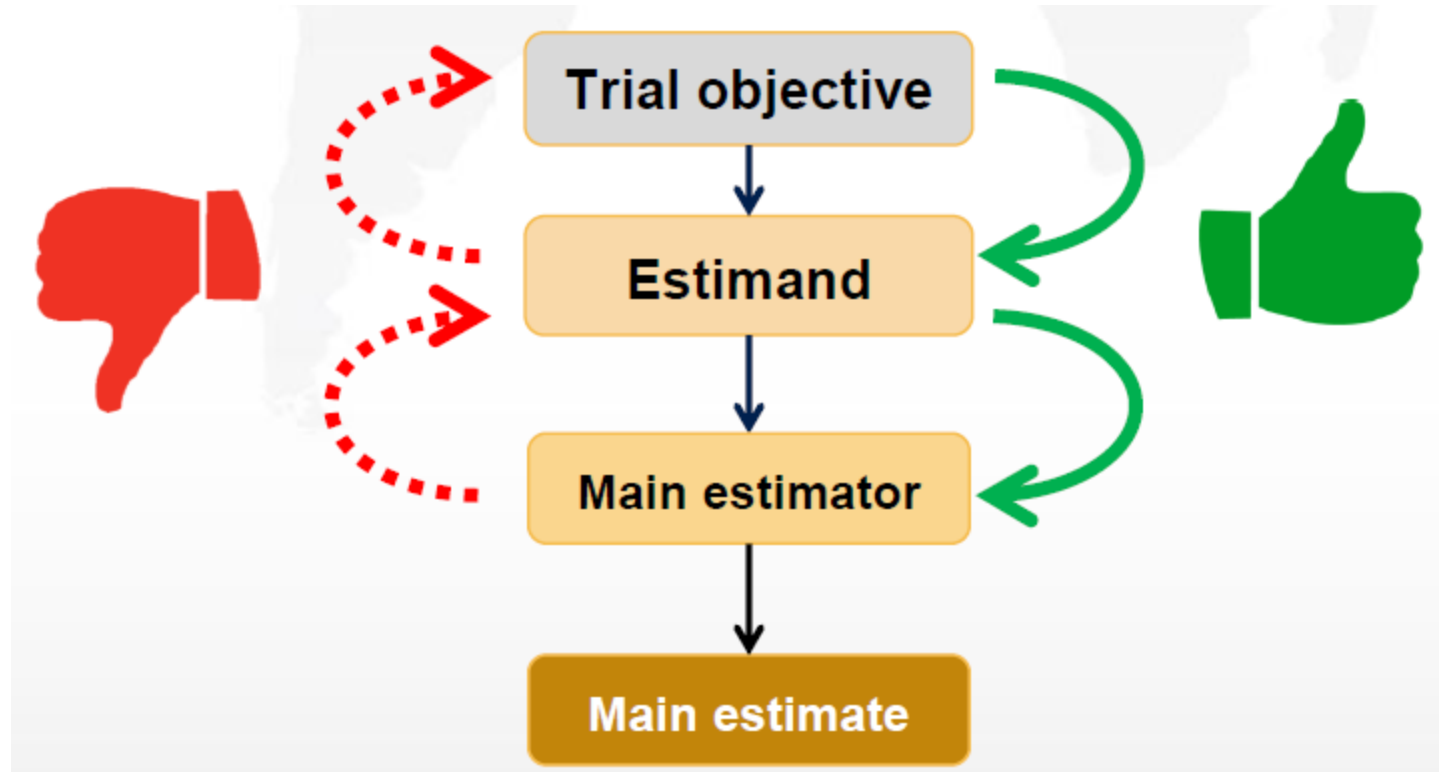
* 第39回ICH即時報告会資料より

新しいフレームワーク



HOW(どのように)からWHAT(何を)に重点変化

留意事項



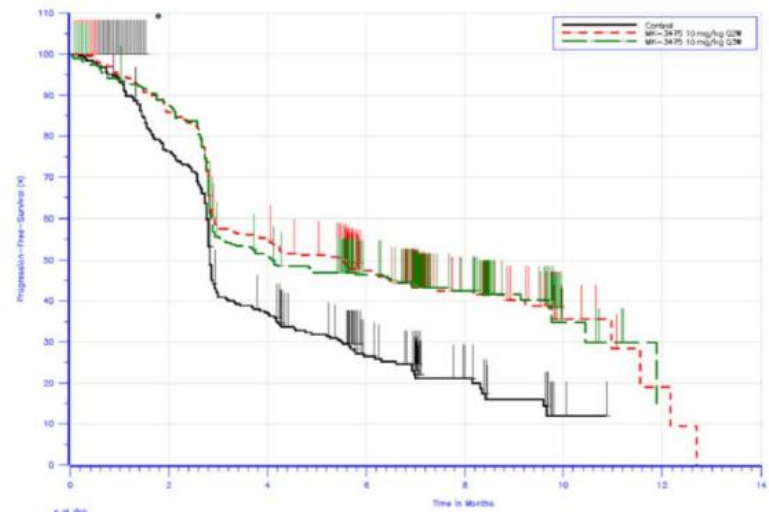
トレーニング資料

- Module 1: Summary
- Module 2: Comprehensive slide deck
 - Module 2.1: Introduction
 - Module 2.2: Framework
 - Module 2.3: Estimands
 - Module 2.4. Impact on Trial design and Conduct
 - Module 2.5. Impact on Trial Analysis
 - Module 2.6. Documenting Estimands and Sensitivity Analysis
- Module 3: Generic Example
 - **A Thinking Process**

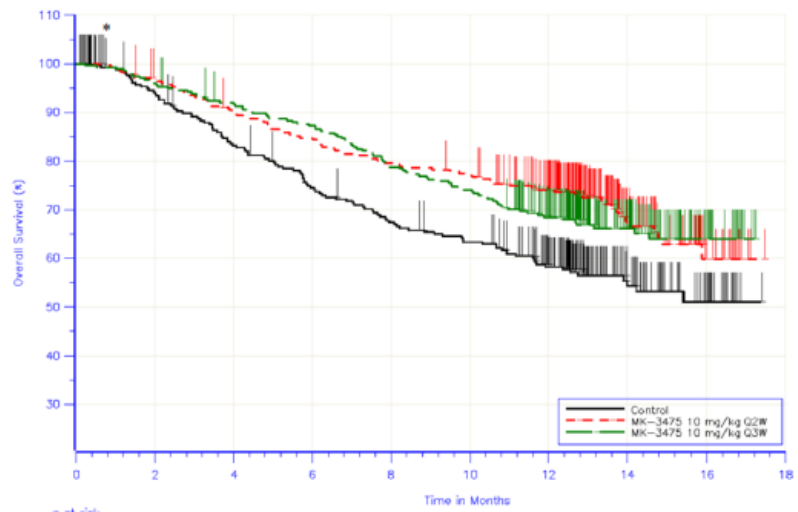
Estimandを構成し、主たる推定量・感度分析などを決めるために考える過程が書かれている

Pembrolizumab Phase 3 study (KEYNOTE-006 study)

- 対象患者: 化学療法歴のない又はipilimumabを含まない1レジメンまでの化学療法歴を有する根治切除不能な悪性黒色腫
- 主要評価項目: 中央判定のPFSとOS



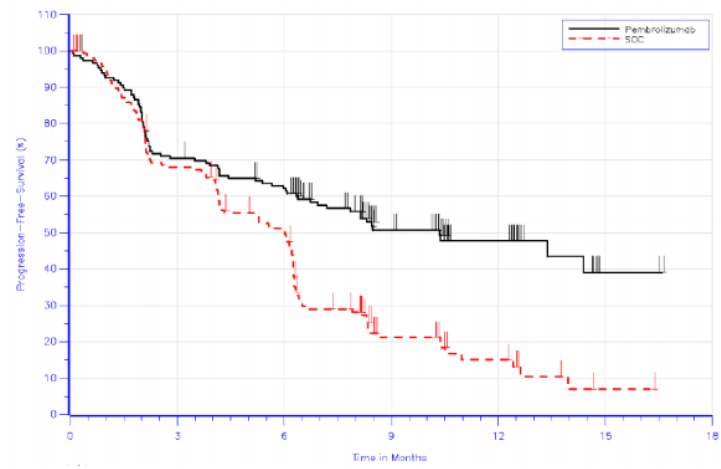
PFS	10 mg/kg Q3W群	10 mg/kg Q2W群	Ipilimumab群
例数	277	279	278
イベント数 (%)	157 (56.7)	157 (56.3)	188 (67.6)
中央値 [95% CI] (月)	4.1 [2.9, 6.9]	5.5 [3.4, 6.9]	2.8 [2.8, 2.9]
ハザード比 [95% CI]	0.58 [0.47, 0.72]	0.58 [0.46, 0.72]	
P値 (片側)	< 0.0001	< 0.0001	



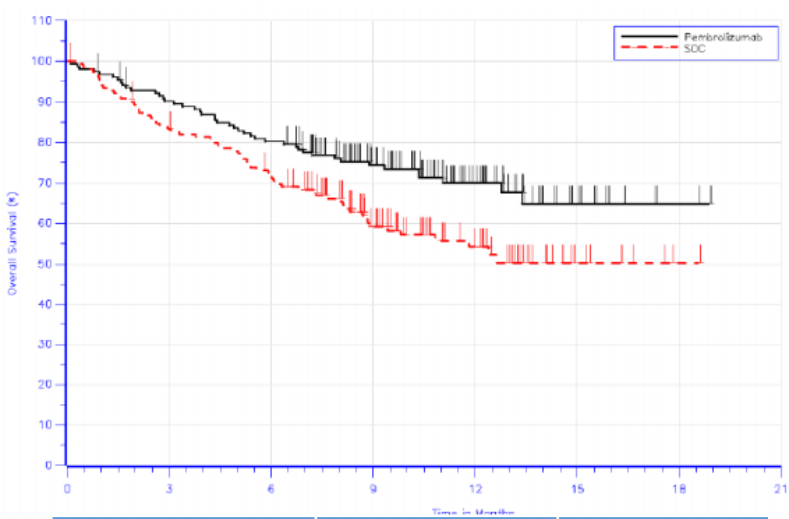
OS	10 mg/kg Q3W群	10 mg/kg Q2W群	Ipilimumab群
例数	277	279	278
イベント数 (%)	92 (33.2)	85 (30.5)	112 (40.3)
中央値 [95% CI] (月)	NE [NE, NE]	NE [NE, NE]	NE [12.7, NE]
ハザード比 [95% CI]	0.69 [0.52, 0.90]	0.63 [0.47, 0.83]	
P値 (片側)	0.00358	0.00052	

Pembrolizumab Phase 3 study (KEYNOTE-024 study)

- 対象患者
化学療法歴のない、EGFR遺伝子変異陰性、ALK融合遺伝子陰性、PD-L1陽性の進行・再発の非小細胞肺癌患者
- 主要評価項目：中央判定のPFS



PFS	Pembrolizumab 群	SOC 群
例数	154	151
イベント数(%)	73 (47.4)	116 (76.8)
中央値 [95% CI] (月)	10.3 [6.7, NE]	6.0 [4.2, 6.2]
ハザード比 [95% CI]	0.50 [0.37, 0.68]	
P値 (片側)	< 0.001	



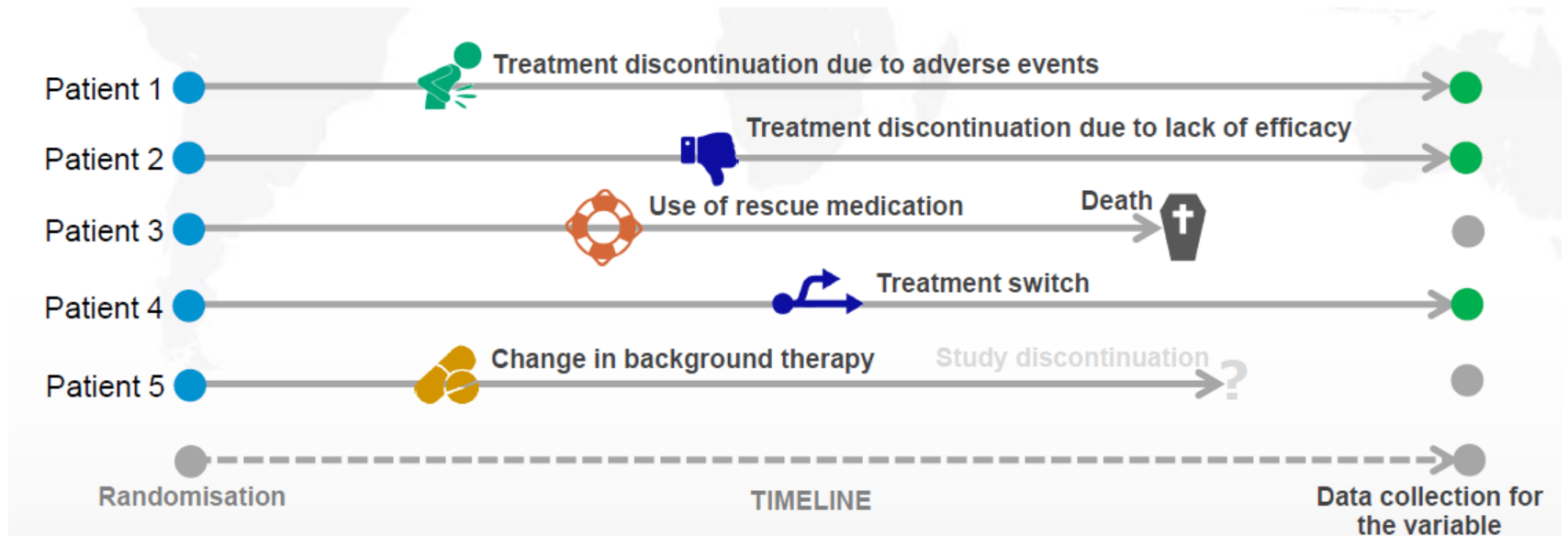
OS	Pembrolizumab 群	SOC 群
例数	154	151
イベント数(%)	44 (28.6)	64 (42.2)
中央値 [95% CI] (月)	NE [NE, NE]	NE [9.4, NE]
ハザード比 [95% CI]	0.60 [0.41, 0.89]	
P値 (片側)	0.005	

A Thinking Process

1. 試験目的を決定づける疾患領域及び治療特性
2. **中間事象の特定**
3. 中間事象に対応するためのストラテジー
4. Estimandの構成
5. 試験デザイン, データ収集, 推定方法の整合性
6. 主解析の仮定とこれらの仮定を検討するための感度解析の特定
7. 選択されたestimandの文書化

2. 中間事象の特定

- 治療開始後に発現し、変数を観測できなくする、又は変数の解釈に影響を与える事象
 - 欠測との違い
(同じ事象でも中間事象になったりならなかったり)
 - 理由等も考慮すべき



議題1

- どのような中間事象が、どのような理由で発生すると想定されるでしょうか？
 - 治療不遵守（服薬計画を遵守できない）
 - 代替治療（レスキュー、併用禁止薬、併用薬変更、増量）
 - 治療の中止（有害事象、有効性不足）
 - 後続ラインの治療
 - switching
 - 死亡
 - 観察不遵守
 - withdrawal from follow-up、転院

A Thinking Process

1. 試験目的を決定づける疾患領域及び治療特性
2. 中間事象の特定
3. **中間事象に対応するためのストラテジー**
4. Estimandの構成
5. 試験デザイン, データ収集, 推定方法の整合性
6. 主解析の仮定とこれらの仮定を検討するための感度解析の特定
7. 選択されたestimandの文書化

3.中間事象に対応するための戦略

- それぞれの中間事象に対し、**関心のある科学的疑問を反映させる戦略**を選択
 - 治療方針戦略
 - 複合戦略
 - 仮想戦略
 - 主要層戦略
 - 治療下戦略
 - ...など

議題2

- 特定された中間事象について、どのような取扱いが想定されるでしょうか？
 - 治療方針ストラテジー：policyとして捉え、問題としない
 - 複合ストラテジー：変数の定義に含める
 - 仮想ストラテジー：発現しなかった状況を想定する
 - 主要層ストラテジー：発現しない症例を関心の対象とする
 - 治療下ストラテジー：発現するまでの反応を関心の対象とする

A Thinking Process

1. 試験目的を決定づける疾患領域及び治療特性
2. 中間事象の特定
3. 中間事象に対応するためのストラテジー
4. Estimandの構成
5. 試験デザイン, データ収集, 推定方法の整合性
6. 主解析の仮定とこれらの仮定を検討するための感度解析の特定
7. 選択されたestimandの文書化

議題3

Estimandを構成する要素

これまでの議論をまとめて、estimandを構成してください

- A: 対象集団
 - 科学的疑問の対象となる患者
- B: 変数
- C: 中間事象
 - 関心のある科学的疑問を反映するために、中間事象をどのように考慮するかという説明
- D: 集団レベルでの変数の要約

設定するestimandと推定量

- 試験目的

治療下での作用(薬理学的な効果)の有無を確認したい

- 主とするestimand

要素	内容
対象集団	〇〇癌の患者
変数	
中間事象	
要約	

- 主とする推定量:
- 感度分析における推定量:
- データ収集で留意すべき事:
- 変数、中間事象の観察をどこまで継続するべきか?

時間があれば、
推定量や感度
分析などにつ
いて議論して
みて下さい

グループディスカッション

- 各テーブルでのグループディスカッションでは、ファシリテーターの方が調整役として進行をお願い致します
- 各グループにて、担当を決めて下さい
 - 書記: PCで議論結果を記録
 - タイムキーパー: 進行時間の確認
 - 発表者: 終了後の結果発表
- グループディスカッション終了後、発表者の方から、3分で結果発表頂きます

第6回 データサイエンス ラウンドテーブル会議

Estimand
～テーブルディスカッション～

用意した6つのケース

疾患領域	ケース	試験の相	主要評価項目	
自己免疫疾患	①	検証的試験(P3試験)	BVAS	- 関心のある指標は寛解。ステロイドも影響 - ステロイド量が減ることにも意義がある
糖尿病	②	検証的試験(P3試験)	HbA1c (52w)	- 非劣性試験 - EMA疾患ガイドラインを参照
慢性腎臓病	③	検証的試験(P3試験)	複合イベント	複合エンドポイント(腎疾患進行及びCV死亡)のイベント試験
がん	④	探索的試験(P2試験)	PFS	- 開発初期 - 目的は腫瘍増悪を抑えることでPOCを説明すること
	⑤	検証的試験(P3試験)	OS	標準治療を対照としたP3試験
	⑥	検証的試験(P3試験)	PFS	⑤のケースに対して、本来評価されるべきはOSだが、代替としてPFSを評価することが許容される

- 各テーブルでいずれかを選択ください(できれば2つ以上)
- 議論がしにくい場合は、ケースを一部変更・追加してください
- 用意した以外のケースについても議論したい場合も、自由に議論ください

グループディスカッション

事例検討：がん以外（ケース①～③）

モデル事例:ニューカラ皮下注(メポリズマブ)のEGPAへの開発
Wechsler et al. NEJM 2017, 376(20): 1921—1932.

グループディスカッション

ケース①: 好酸球性多発血管炎性肉芽腫症

ケース①: EGPA

- 先行するアレルギー兆候(気管支喘息, アレルギー性鼻炎, 好酸球性副鼻腔炎等)と末梢血中の好酸球数増加を伴い, 組織学的には周囲組織への好酸球数浸潤を伴う細小血管の肉芽腫又はファブリノイド壊死性血管炎を呈するANCA関連血管炎の一病型
- 血管炎発症時の主な臨床症状は, 発熱, 体重減少, 多発性単神経炎, 消化管出血, 紫斑, 多関節痛(炎), 筋肉痛, 筋力低下等であり, 一部の患者では, 血管炎発症1~3週間で致死的な虚血による消化管障害(消化管穿孔, イレウス), 心障害(不整脈, 心不全, 心停止)を呈する.
- 国内における患者数は2008年時点で1866名と推計.

ケース①: EGPAに対する標準的療法

- 経口ステロイド剤が使用される
- 治療目標達成後は速やかに経口ステロイドを減量し、必要最小用量での維持、可能であれば投与中止が望まれる
- EGPAはしばしば再燃を繰り返す慢性の病態を呈し、経口ステロイドの減量に難渋することも少なくない

ケース①:想定している状況

- 検証的試験
- 目的は本剤の有効性及び安全性の検討
- プラセボ対照ランダム化二重盲検比較試験
 - プラセボ群, 開発薬剤群の2群
 - 4週間隔で52週間皮下投与し, その後, 8週間の後観察

ケース①:試験目的と科学的疑問

- 本試験の目的は、日常診療に即した条件の下で、本剤を経口ステロイドでも寛解しきらない患者に対して使用し、寛解に至らせるとともに、経口ステロイドを減量し、病状をコントロールすることができるかどうかを評価すること
- 関心のある科学的疑問
 - 本剤により経口ステロイドでも寛解しきれない患者を寛解に至らせることができるのか
 - 本剤により経口ステロイドがどれだけ減量維持できるのか
 - 本剤を中止した後、経口ステロイドの量及び病状がリバウンドしないか

ケース①:試験の対象患者

- 再燃の既往がある又は難治性の経口ステロイド投与下のEGPA患者が対象
 - (1) 2年以内かつ12週間以上前に, 7.5mg以上の経口ステロイド剤投与下で再燃の既往がある.
 - (2) 寛解導入のための治療を3カ月以上施行しても寛解に至らなかった, 又は6カ月以内の経口ステロイド漸減中, 7.5mg/日以上 of 経口ステロイド投与下で再燃した.

ケース①:試験デザインの詳細

- 開発薬剤又はプラセボを4週間隔で52週間投与
 - 経口ステロイドは、ベースラインの4週以上前から一定用量を併用することとされ、ベースラインから4週後以降、治験担当医師の判断で漸減する計画
 - 規定されている範囲内で増量することも可
- 52週の治験薬投与後、8週間の後観察

ケース① : BVAS

(Birmingham vasculitis activity score)

- 全身性血管炎の疾患活動性を評価する指標
- 全身及び8つの臓器における56症状/兆候を点数化し, 臓器別の点数を加算した総点数
- スコアの範囲は0～63
- スコアが高いほど疾患活動性が高い.
- BVAS=0は疾患活動性が消失した状態

ケース①：ディスカッションポイント

- 寛解だけではなく、経口ステロイドの減量が出来たことも治療目標であり、当該目標をどう考慮するかについても考えながら議論してください

グループディスカッション

ケース②: 糖尿病 非劣性試験

ケース②: 2型糖尿病とは

- 2型糖尿病薬の開発
 - 運動療法や食事療法から開始し、血糖コントロールできない人が薬物療法を開始。運動・食事療法等は引き続き行う。初期は単剤から開始し、コントロールできない場合は併用・切り替えを行う。ほとんどの患者は併用(2剤/3剤)で使用する。
 - 第一選択薬は安全で安価
 - ヨーロッパでは、保険償還の観点から、実薬対照試験も重要と考えられている
 - 評価項目のHbA1cの変化量は確立した指標

ケース②：想定している状況

モデル事例：EMA疾患ガイドライン

- 想定している試験
 - 対象患者：第一選択薬で効果不十分な糖尿病患者
 - 試験デザイン：国際共同実薬対照非劣性試験
 - 新薬群：標準薬（第一選択薬）＋新薬（新しい作業機序）
 - 対照群：標準薬（第一選択薬）＋他剤X
 - 試験期間：52週間
 - 評価項目：52週間後のHbA1Cの変化量

ケース②：想定している状況

モデル事例：EMA疾患ガイドライン

- 科学的疑問
 - 第一選択薬と併用した時の、それ以外の薬剤の影響を除いた他剤Xに対する新薬の効果
- 想定している状況
 - 糖尿病は長期間投与される事が想定されており、長期の安全性及び有効性の確立が重要
 - 基本的に第一選択薬の用量変更や他剤の追加は行わないこと。ただし長期投与試験でもあり倫理的な理由等やむを得ない場合は可
 - 主な副作用：低血糖（重度な場合はインスリン注射）
 - 薬剤を中止しても、継続ほどではないがある程度効果は持続
 - 試験中止率：両群とも約20%
 - 新薬群：有害事象5%，有効性欠如5%，その他（同意撤回等）10%
 - 対照群：有害事象5%，有効性欠如10%，その他（同意撤回等）5%
 - 死亡率：両群とも1%

ケース②：ディスカッションポイント

モデル事例：EMA疾患ガイドライン

- 非劣性試験における懸念される事, 考慮すべき事があるかも含め議論してください
- 今まで感度分析として実施してきたPPS解析について議論してください

グループディスカッション

ケース③: 慢性腎臓病 イベント試験

ケース③: 慢性腎臓病

- 慢性腎臓病
 - 高い有病率
 - 原因は様々
 - 原疾患(糖尿病, 高血圧, 脂質異常症など)
 - 加齢など
- 治療の目的(腎機能は一旦悪化すると改善しない。機能を維持させることが重要)
 - 腎機能が悪化し末期腎不全になるのを防ぐ
 - 合併症(心筋梗塞, 脳卒中)を防ぐ
- 軽度でもCVイベントを起こし, 末期腎不全に至る前に死亡する場合もある。末期腎不全になると, 透析や腎移植が必要となるから避けたい

ケース③：想定している状況

モデル事例

ID=NCT03594110 from Clinical trials.gov

- 想定している薬剤の背景
 - 経口糖尿病治療薬
 - CVイベント発症リスクが高い糖尿病患者を対象に, CVイベントをoutcomeとする試験を実施
 - 腎複合イベントに対する事後解析で良好な結果
 - 作用機序からも腎保護作用効果は考えられうる
- ⇒ 糖尿病患者以外も含め, 慢性腎疾患の患者を対象に, 新たに試験実施

ケース③：想定している状況

モデル事例

ID=NCT03594110 from Clinical trials.gov

- 想定している試験
 - 対象患者：腎疾患進行のリスクのある慢性腎臓病患者
 - 標準的な慢性腎臓病の治療を受けている中等度以上の腎機能もしくはアルブミン尿の高い患者
 - 原疾患：問わない
 - 試験デザイン：国際共同プラセボ対照add-on試験。イベント試験
 - 試験期間：3.5年（中央値3年）
 - 評価項目：複合エンドポイント
 - 腎疾患進行（透析や腎移植が必要となる状態 or 腎関連死亡 or 40%以上の腎機能減少）
 - CV死亡

ケース③：想定している状況

モデル事例

ID=NCT03594110 from Clinical trials.gov

- 科学的疑問
 - 長期使用における臨床的重要なアウトカムに対する治療効果
- 想定している状況
 - 長期の有効性及び安全性の確立が重要
 - 海外での保険適用には、実臨床下に近いデータが要求
 - 試験完了率
 - 実薬群：97%，プラセボ群97%
 - CV死亡率
 - 実薬群：4%，プラセボ群6%
 - 死亡率（理由を問わない）
 - 実薬群：6%，プラセボ群8.5%
 - 試験薬中止率
 - 実薬群：20%，プラセボ群25%

グループディスカッション

事例検討：がん（ケース④～⑥）

テーマ：がん

- 対象疾患：がん
- 論点
 - がん患者を対象とした臨床試験のestimand（特に、中間事象）を整理したい
- 3つのケースを用意しました

Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics - Guidance for Industry, FDA 2018

Endpoint	Advantages	Disadvantages
Overall Survival (OS)	• Easily and precisely measured • Generally based on objective and quantitative assessment	• May be affected by switch-over of control to treatment or subsequent therapies • Needs longer follow-up • Includes noncancer deaths
Symptom Endpoints (patient reported outcomes)	• Generally assessed earlier and with smaller sample size compared with survival studies	• Blinding is important for assessing the endpoint • Potentially subject to assessment bias, particularly in open-label studies • Lack of validated instruments in many disease areas • Definitions vary among studies • Balanced timing of assessments among treatment arms is critical

Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics - Guidance for Industry, FDA 2018

Endpoint	Advantages	Disadvantages
Disease-Free Survival or Event-Free Survival	• Generally assessed earlier and with smaller sample size compared with survival studies • Generally based on objective and quantitative assessment	• Potentially subject to assessment bias, particularly in open-label studies • Definitions vary among studies • Balanced timing of assessments among treatment arms is critical • Includes noncancer deaths
Objective Response Rate (ORR) 薬の抗腫瘍効果を測る 直接的な指標	• Generally assessed earlier and with smaller sample size compared with survival studies • Effect on tumor attributable to drug(s), not natural history • Generally based on objective and quantitative assessment	• Definitions vary among studies • Frequent radiological or other assessments • May not always correlate with survival

Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics - Guidance for Industry, FDA 2018

Endpoint	Advantages	Disadvantages
Complete Response (CR)	• Generally assessed earlier and with smaller sample size compared with survival studies • Effect on tumor attributable to drug(s), not natural history • Generally based on objective and quantitative assessment	• Definitions vary among studies • Frequent radiological or other assessments • May not always correlate with survival
Progression Free Survival (PFS) or Time to Progression (TTP) Natural historyを反映する time-to-eventタイプの指標	• Generally assessed earlier and with smaller sample size compared with survival studies • Measurement of stable disease included • Generally based on objective and quantitative assessment	• Potentially subject to assessment bias, particularly in open-label studies • Definitions vary among studies • Frequent radiological or other assessments • Balanced timing of assessments among treatment arms is critical • May not always correlate with survival

RECIST ガイドライン

固形がんの治療効果判定のための新ガイドライン

- ベースライン時に「標的病変」と「非標的病変」を特定し、各病変の径和を測定して治療開始
- 規定したスケジュールに沿って、治療開始後の各時点の抗腫瘍効果を判定

標的病変の効果判定

効果	定義
CR	すべての標的病変の消失 すべてのリンパ節病変は短径10mm未満に縮小
PR	ベースライン径和から標的病変が30%以上減少
PD	経過中の最小の径和と比較して標的病変が20%以上増加、かつ径和が5mm以上増加
SD	PRでなくPDでない

非標的病変の効果判定

効果	定義
CR	すべての非標的病変の消失、かつ腫瘍マーカー値が基準値上限以下。すべてのリンパ節病変は短径10mm未満に縮小
Non-CR/ non-PD	1つ以上の非標的病変の残存かつ/または腫瘍マーカー値が基準値上限を超える
PD	明らかな増悪

CR (complete response; 完全奏効), PR (partial response; 部分奏効), PD (progressive disease; 進行), SD (stable disease; 安定)

RECIST ガイドライン(続き)

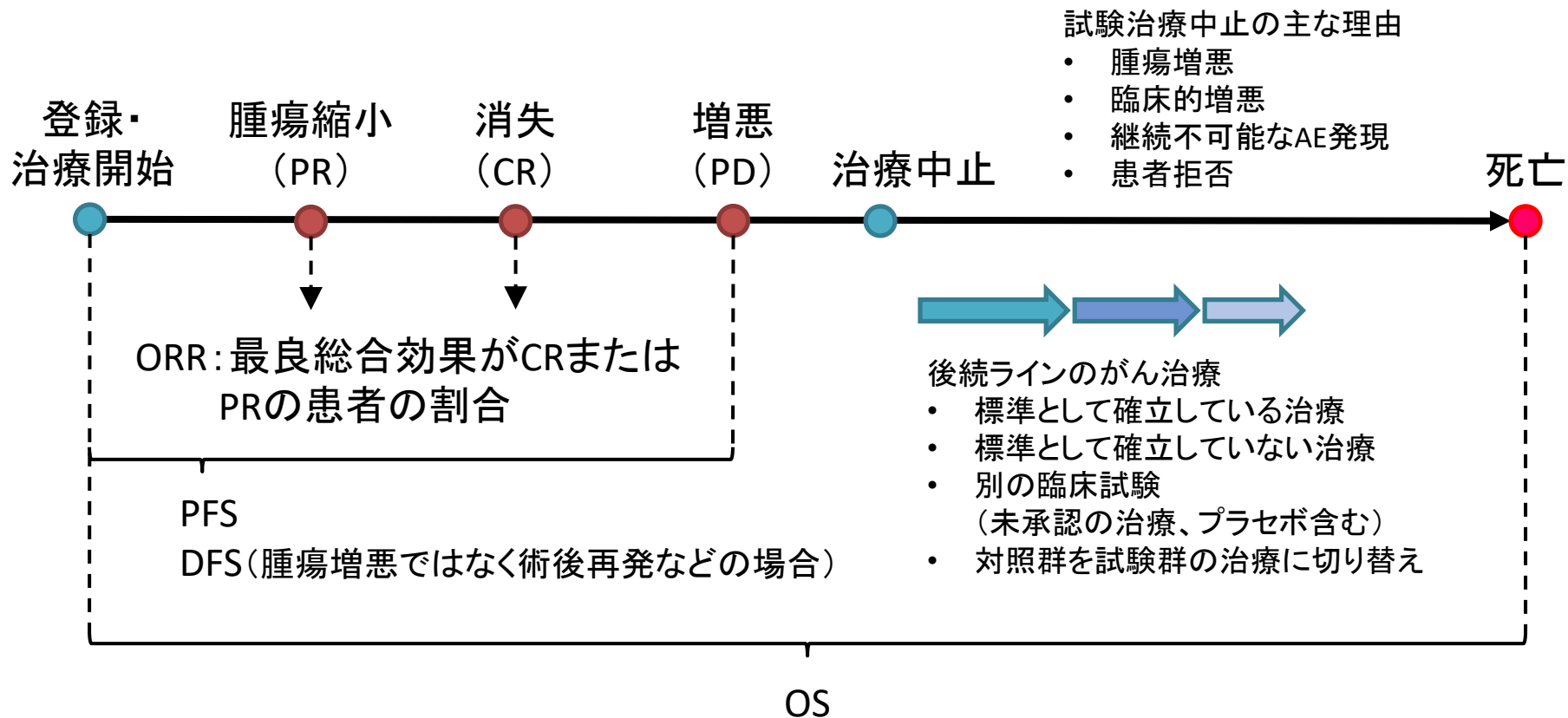
固形がんの治療効果判定のための新ガイドライン

- 各時点の標的病変と非標的病変の抗腫瘍効果、新病変の有無によって、総合効果を判定

標的病変	非標的病変	新病変	総合効果
CR	CR	なし	CR
CR	Non-CR/non-PD	なし	PR
CR	評価なし	なし	PR
PR	Non-PD or 評価欠損あり	なし	PR
SD	Non-PD or 評価欠損あり	なし	SD
評価欠損あり	Non-PD	なし	NE
PD	問わない	あり or なし	PD
問わない	PD	あり or なし	PD
問わない	問わない	あり	PD

- すべての時点の総合効果から、最も良い判定結果として、最良総合効果を決定

試験治療の推移と評価項目

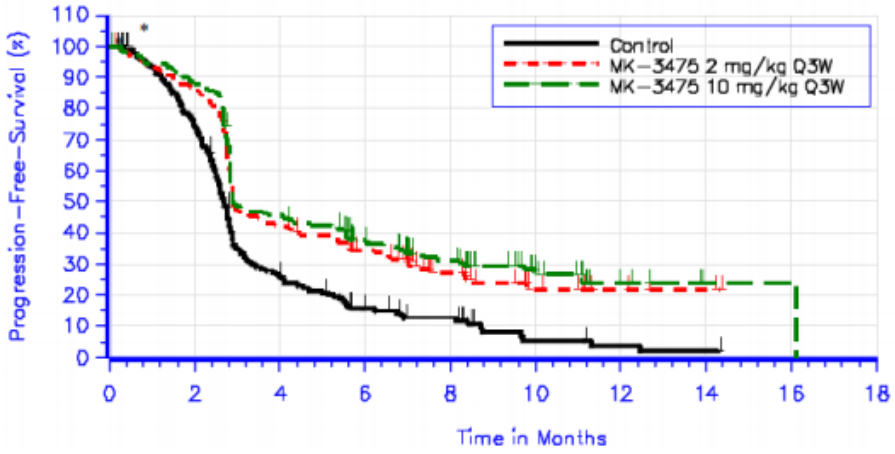


3つのケース

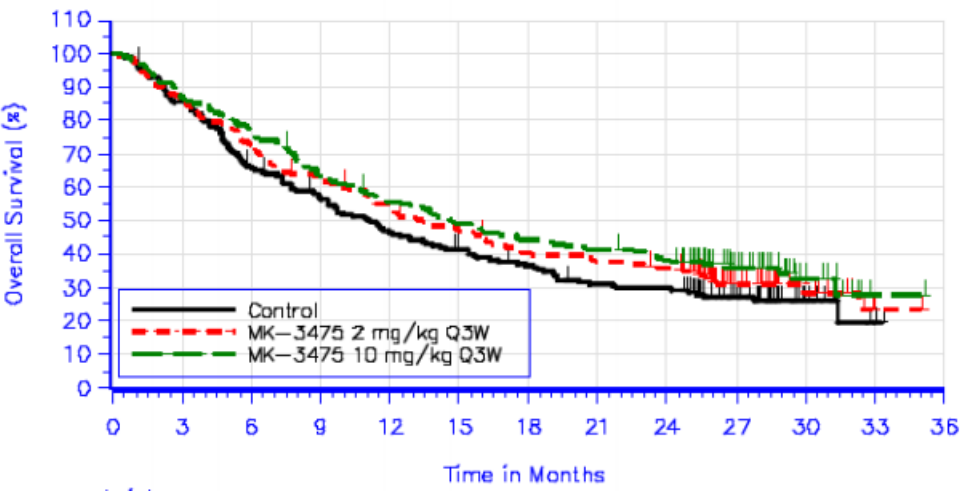
ケース	試験の相	主要評価項目	治療の背景、意図
④	探索的試験 (P2試験)	PFS	• 同時対照に用量群又はプラセボを設定した比較試験 • 開発初期で、POCはまだ証明されていない • 腫瘍縮小を目標とした薬ではないためORRでは薬剤効果を解釈することができない。PFSで説明したい
⑤	探索的試験 (P3試験)	OS	• 同時対照に標準治療を設定したランダム化比較試験 • 用量設定とPOCは確認済み。標準治療を上回ることを検証したい
⑥	探索的試験 (P3試験)	PFS	• 同時対照に標準治療を設定したランダム化比較試験 • 用量設定とPOCは確認済み。標準治療を上回ることを検証したい • 本来評価されるべきはOSだが、代替としてPFSが許容される

Pembrolizumab Phase 2 study (KEYNOTE-002 study)

- 対象患者：治療歴を有する根治切除不能な悪性黒色腫
- 主要評価項目：中央判定のPFSとOS



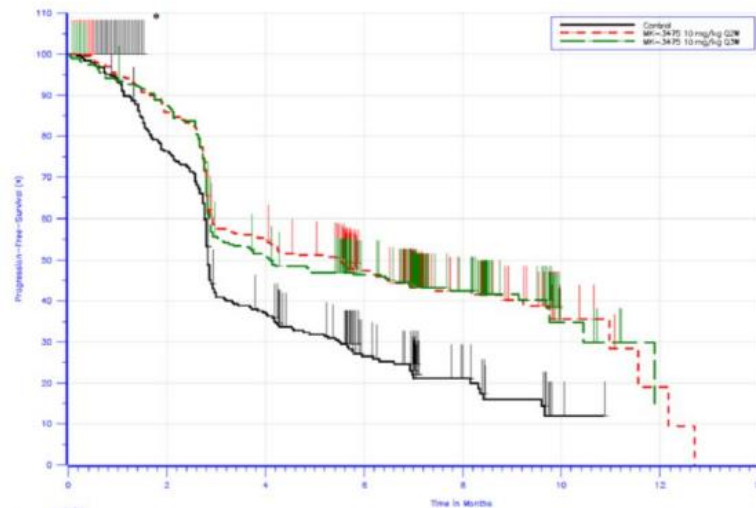
PFS	2 mg/kg 群	10 mg/kg 群	ICC(Control) 群
例数	180	181	179
イベント数 (%)	129 (71.7)	126 (69.6)	155 (86.6)
中央値 [95% CI] (月)	2.9 [2.8, 3.8]	2.9 [2.8, 4.7]	2.7 [2.5, 2.8]
ハザード比 [95% CI]	0.57 [0.45, 0.73]	0.50 [0.39, 0.64]	
P値 (片側)	< 0.0001	< 0.0001	



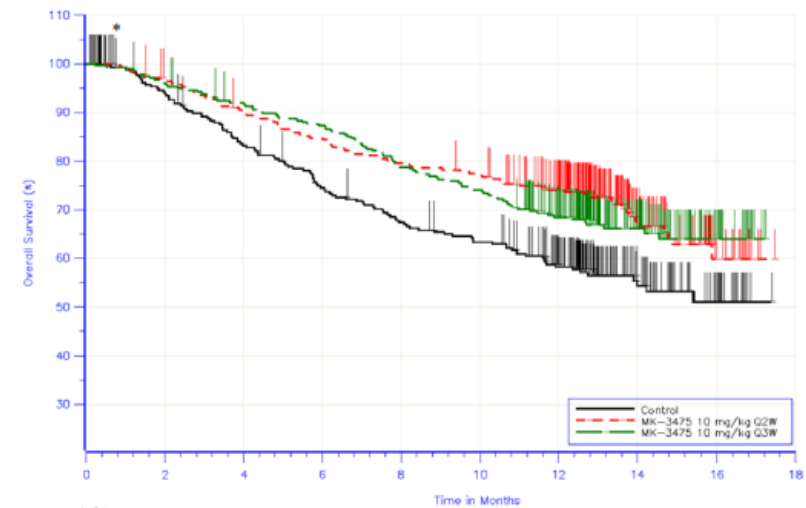
OS	2 mg/kg 群	10 mg/kg 群	ICC(Control) 群
例数	180	181	179
イベント数 (%)	123 (68.3)	69 (38.1)	78 (43.6)
中央値 [95% CI] (月)	13.4 [10.2, NE]	12.5 [9.7, NE]	11.6 [9.0, 16.3]
ハザード比 [95% CI]	0.88 [0.64, 1.22]	0.78 [0.56, 1.08]	
P値 (片側)	0.2294	0.0664	

Pembrolizumab Phase 3 study (KEYNOTE-006 study)

- 対象患者: 化学療法歴のない又はipilimumabを含まない1レジメンまでの化学療法歴を有する根治切除不能な悪性黒色腫
- 主要評価項目: 中央判定のPFSとOS



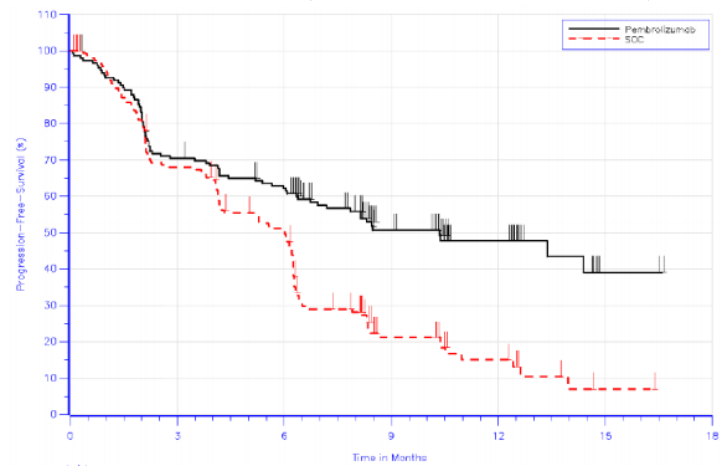
PFS	10 mg/kg Q3W群	10 mg/kg Q2W群	Ipilimumab群
例数	277	279	278
イベント数 (%)	157 (56.7)	157 (56.3)	188 (67.6)
中央値 [95% CI] (月)	4.1 [2.9, 6.9]	5.5 [3.4, 6.9]	2.8 [2.8, 2.9]
ハザード比 [95% CI]	0.58 [0.47, 0.72]	0.58 [0.46, 0.72]	
P値 (片側)	< 0.0001	< 0.0001	



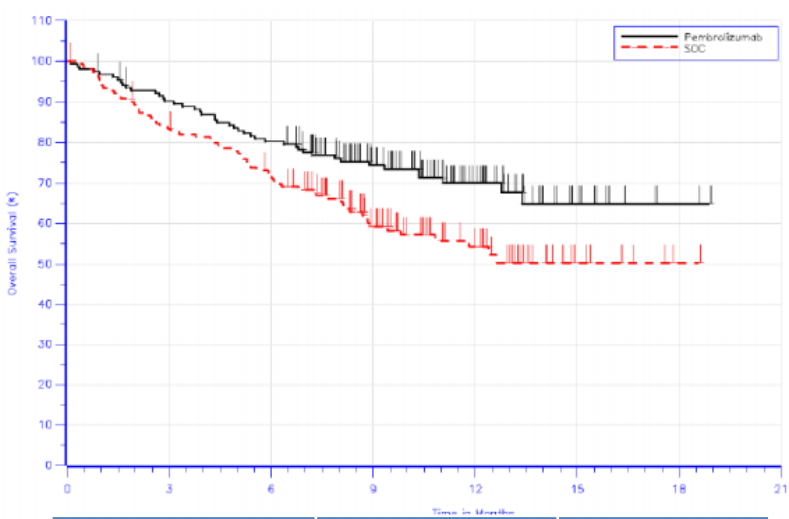
OS	10 mg/kg Q3W群	10 mg/kg Q2W群	Ipilimumab群
例数	277	279	278
イベント数 (%)	92 (33.2)	85 (30.5)	112 (40.3)
中央値 [95% CI] (月)	NE [NE, NE]	NE [NE, NE]	NE [12.7, NE]
ハザード比 [95% CI]	0.69 [0.52, 0.90]	0.63 [0.47, 0.83]	
P値 (片側)	0.00358	0.00052	

Pembrolizumab Phase 3 study (KEYNOTE-024 study)

- 対象患者
化学療法歴のない、EGFR遺伝子変異陰性、ALK融合遺伝子陰性、PD-L1陽性の進行・再発の非小細胞肺癌患者
- 主要評価項目：中央判定のPFS



PFS	Pembrolizumab 群	SOC 群
例数	154	151
イベント数(%)	73 (47.4)	116 (76.8)
中央値 [95% CI] (月)	10.3 [6.7, NE]	6.0 [4.2, 6.2]
ハザード比 [95% CI]	0.50 [0.37, 0.68]	
P値 (片側)	< 0.001	



OS	Pembrolizumab 群	SOC 群
例数	154	151
イベント数(%)	44 (28.6)	64 (42.2)
中央値 [95% CI] (月)	NE [NE, NE]	NE [9.4, NE]
ハザード比 [95% CI]	0.60 [0.41, 0.89]	
P値 (片側)	0.005	

PFS解析の打ち切りルール

KEYNOTE-002, 006, 024 studyで共通

Situation	Primary Analysis	Sensitivity Analysis 1	Sensitivity Analysis 2
No PD and no death; new anticancer treatment is not initiated	Censored at last disease assessment	Censored at last disease assessment	Censored at last disease assessment if still on study therapy; progressed at treatment discontinuation otherwise
No PD and no death; new anticancer treatment is initiated	Censored at last disease assessment before new anticancer treatment	Censored at last disease assessment before new anticancer treatment	Progressed at date of new anticancer treatment
PD or death documented after ≤ 1 missed disease assessment	Progressed at date of documented PD or death	Progressed at date of documented PD or death	Progressed at date of documented PD or death
PD or death documented after ≥ 2 missed disease assessments	Progressed at date of documented PD or death	Censored at last disease assessment prior to the ≥ 2 missed disease assessment	Progressed at date of documented PD or death

KEYNOTE-002 & 006 studyの特徴

Phase 2 study (KEYNOTE-002 study)

- 標準治療が確立してない患者を対象としているため、対照群にICC (investigator choice chemotherapies) を設定
- ICC群に登録され、基準*を満たした患者は12週目以降に実薬投与に切り替えを許容
 - 実際には、ICC群の86/179 (48.0%) に実薬が投与された

* ①PS 0 又は 1、②新たな脳転移の出現又は症状・兆候を認めないこと、③化学療法による有害事象が Grade 1 以下に回復していること、④中央画像判定で疾患進行が確認されていること

Phase 3 study (KEYNOTE-006 study)

- 治療GLにて報告されていたipilimumabを対照群に設定
- Ipilimumab治療 → pembrolizumab治療の切り替えはなし

KEYNOTE-024 studyの特徴

- 治療GLにて報告されていたSOC(CDDP/GEM等)を対照群に設定
- PFSを主要評価項目に設定したことについて

スポンサー

疾患進行に伴う臨床症状の悪化を遅らせることに臨床的意義があると考えられる

PMDA

延命を期待して施行されるものであり、主要評価項目はOSを設定すべきであった

設定すべき評価項目というより、estimandが違う

論点 — ケース④

- 薬剤コンセプト
 - 腫瘍縮小するかどうかは分からない
 - 腫瘍安定を維持できれば、薬剤効果を説明できる
- 試験目的
 - まだPOCは証明されていない
 - 治療下での作用(薬理学的な効果)の有無を確認したい
 - natural historyと薬剤効果を分けるため、比較試験にて説明したい

論点 — ケース⑤

- 同時対照に標準治療を設定したランダム化比較試験
- 試験目的
 - 用量設定とPOCは確認済み
 - 延命効果について、標準治療を上回ることを検証したい

論点 — ケース⑥

- 同時対照に標準治療を設定したランダム化比較試験
- 試験目的
 - 用量設定とPOCは確認済み。標準治療を上回ることを検証したい
 - 本来評価されるべきは延命効果だが、代替としてPFSが許容される(とスポンサーは信じている)

追加の論点（もし時間があれば）

Estimandは異なるか？	ケース④	ケース⑤	ケース⑥
対象集団のがんの種類の違い （例：胃癌と肺癌）			
対象集団の治療ラインの違い （例：一次治療、標準治療無効、 術前・術後）			
薬の作用機序（殺細胞、分子標 的、抗体、がん免疫）の違い			

参考資料

- Ribas A et al. Pembrolizumab versus investigator-choice chemotherapy for ipilimumab-refractory melanoma (KEYNOTE-002): a randomised, controlled, phase 2 trial. Lancet Oncol 2015;16:908-918.
- Robert C et al. Pembrolizumab versus Ipilimumab in Advanced Melanoma. N Engl J Med 2015; 372:2521-2532.
- Reck M et al. Pembrolizumab versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung Cancer. N Engl J Med 2016; 375:1823-1833.

結果記録用 テンプレート

どのような中間事象があるか、発生理由別に書き出してみよう
(変数への影響の有無に関わらず)
以下の記載は一例なので、他にあればそれも書き出してみよう

中間事象

中間事象	発生理由	変数を観測できなくする？	変数の解釈に影響する？	実臨床で起こりうる？
試験治療中止	腫瘍増悪			
	臨床的増悪			
	AE発現			
	患者拒否			
代替治療	併用禁止薬			
	後続でがん治療(標準治療)			
	後続でがん治療(標準でない治療)			
	後続でがん治療 (別の臨床試験、プラセボ含む)			
	対照群を試験群の治療に切り替え			
死亡	がんによる死亡			
	がん以外の疾患			
	事故			
	不明			

どのような中間事象があるか、発生理由別に書き出してみましょう
(変数への影響の有無に関わらず)
以下の記載は一例なので、他にあればそれも書き出してみましょう

中間事象

中間事象	発生理由	ストラテジー	判断した理由
試験治療中止	腫瘍増悪		
	臨床的増悪		
	AE発現		
	患者拒否		
代替治療	併用禁止薬		
	後続でがん治療(標準治療)		
	後続でがん治療(標準でない治療)		
	後続でがん治療 (別の臨床試験、プラセボ含む)		
	対照群を試験群の治療に切り替え		
死亡	がんによる死亡		
	がん以外の疾患		
	事故		
	不明		

設定するestimandと推定量

- 主とするestimand

要素	内容
対象集団	〇〇癌の患者
変数	
中間事象	• • • • •
要約	

- 主とする推定量:
- 感度分析における推定量:
- データ収集で留意すべき事:
- 変数、中間事象の観察をどこまで継続するべきか?

Estimand

【第6回データサイエンスラウンドテーブル会議】
2019年3月11日

パネルディスカッション

パネルディスカッション

- 各テーブルで挙げられた論点を踏まえ，総合討論としてパネリストを中心とし参加者全員が参加するパネルディスカッションを実施
 - 事前アンケートに寄せられた話題も考慮した
- 論点を司会から紹介
- パネリストでの議論の後，会場から意見を受け付ける

パネルディスカッションの論点

1. 代替エンドポイント
 - 主要評価項目が代替エンドポイントの場合、特に留意すべきことはあるか？
2. 治療の切り替え
 - 対照群で被験薬に切り替えを許容する場合、各取扱いの長所・短所は？
3. 非劣性試験での考え方
 - 非劣性試験でのPPSの役割は？必要性は？
 - PPSの代わりに主要層として実施することは妥当か？
4. 主要層
 - Dong and Lu (Pharm Stat 2019)を主要層アプローチと言って良いのか？
5. 試験中止
 - 試験中止は、中間事象の定義にあてはまるように見えるが、中間事象ではないのか？

論点1 代替エンドポイント

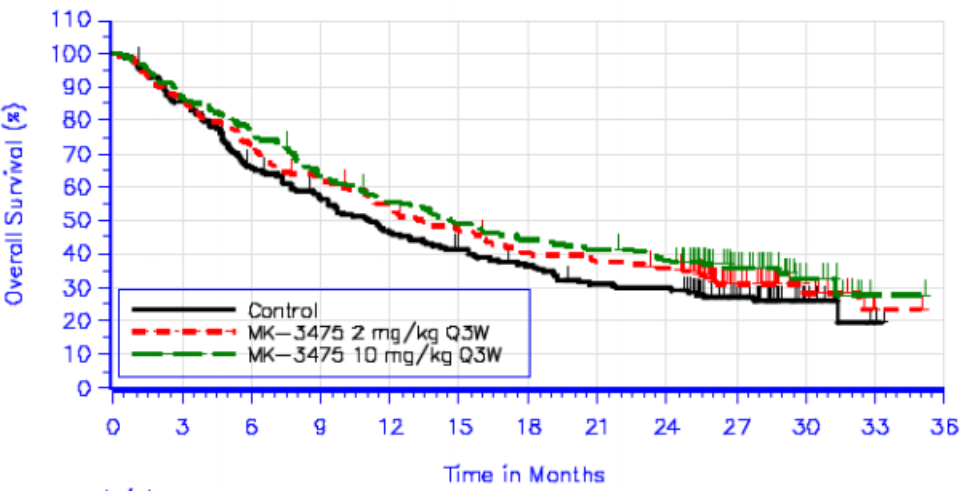
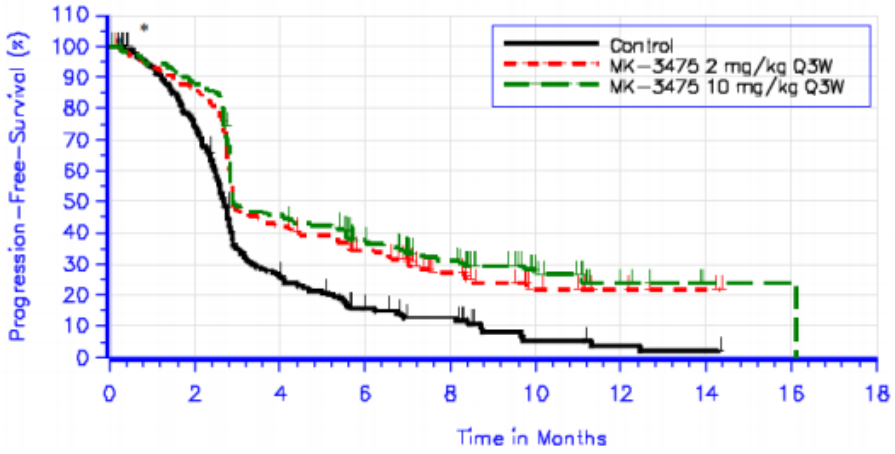
主要評価項目が代替エンドポイントの場合、特に留意すべきことはあるか？

(例)がん臨床試験でのPFS

- 主要評価項目PFSのestimandの構築(例えば要素C、D)でOSを意識すべきか？
- 同じ主要評価項目PFSでも、早期の相でPOCを見る場合と、検証試験で代替エンドポイントとする場合に違いはあるか？
 - KEYNOTE-002 studyとKEYNOTE-024 study

Pembrolizumab Phase 2 study (KEYNOTE-002 study)

- 対象患者：治療歴を有する根治切除不能な悪性黒色腫
- 主要評価項目：中央判定のPFSとOS

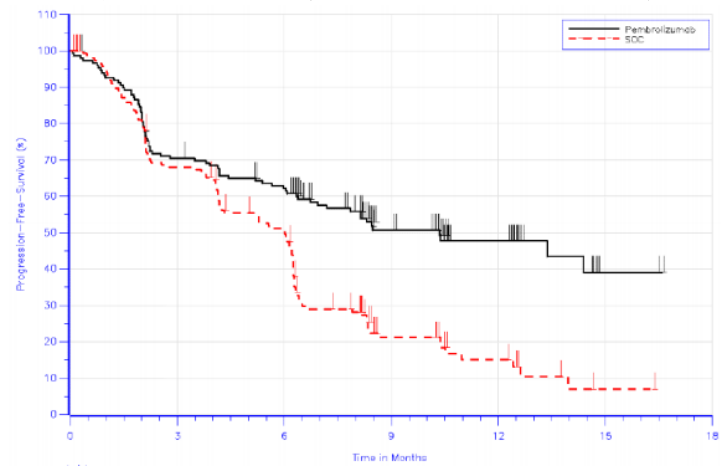


PFS	2 mg/kg 群	10 mg/kg 群	ICC(Control) 群
例数	180	181	179
イベント数 (%)	129 (71.7)	126 (69.6)	155 (86.6)
中央値 [95% CI] (月)	2.9 [2.8, 3.8]	2.9 [2.8, 4.7]	2.7 [2.5, 2.8]
ハザード比 [95% CI]	0.57 [0.45, 0.73]	0.50 [0.39, 0.64]	
P値 (片側)	< 0.0001	< 0.0001	

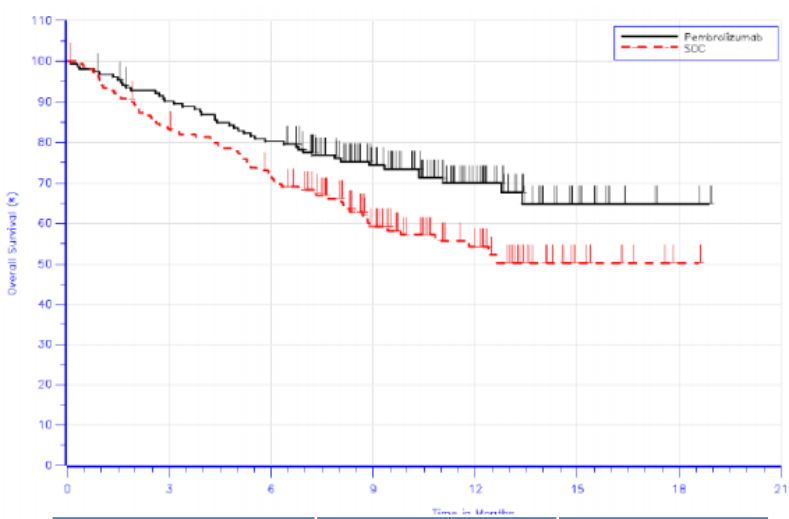
OS	2 mg/kg 群	10 mg/kg 群	ICC(Control) 群
例数	180	181	179
イベント数 (%)	123 (68.3)	69 (38.1)	78 (43.6)
中央値 [95% CI] (月)	13.4 [10.2, NE]	12.5 [9.7, NE]	11.6 [9.0, 16.3]
ハザード比 [95% CI]	0.88 [0.64, 1.22]	0.78 [0.56, 1.08]	
P値 (片側)	0.2294	0.0664	

Pembrolizumab Phase 3 study (KEYNOTE-024 study)

- 対象患者
化学療法歴のない、EGFR遺伝子変異陰性、ALK融合遺伝子陰性、PD-L1陽性の進行・再発の非小細胞肺癌患者
- 主要評価項目：中央判定のPFS



PFS	Pembrolizumab 群	SOC 群
例数	154	151
イベント数(%)	73 (47.4)	116 (76.8)
中央値 [95% CI] (月)	10.3 [6.7, NE]	6.0 [4.2, 6.2]
ハザード比 [95% CI]	0.50 [0.37, 0.68]	
P値 (片側)	< 0.001	



OS	Pembrolizumab 群	SOC 群
例数	154	151
イベント数(%)	44 (28.6)	64 (42.2)
中央値 [95% CI] (月)	NE [NE, NE]	NE [9.4, NE]
ハザード比 [95% CI]	0.60 [0.41, 0.89]	
P値 (片側)	0.005	

論点2 治療の切り替え

対照群で被験薬に切り替えを許容する場合、
各取扱いの長所・短所は？

- 治療方針ストラテジー（切り替え後の成績も含める）
- 仮想ストラテジー（切り替えによる効果を除く）
- 治療下ストラテジー

（例）KEYNOTE-002 study（ケース④）

- 標準治療が確立してないため、対照群はinvestigator choice
- 12週目以降で基準*を満たした患者に実薬投与切り替え許容

* ①PS 0 又は 1、②新たな脳転移の出現又は症状・兆候を認めないこと、③化学療法による有害事象が Grade 1 以下に回復していること、④中央画像判定で疾患進行が確認されていること

論点3 非劣性試験

- Estimandを構成するにあたり優越性試験と非劣性試験で異なる点は何か？
- PPSの利用上の問題点は？
- 主要層ストラテジーは有益か？
 - 使用上の問題点は？

例) ケース② 糖尿病非劣性試験

– EMA糖尿病疾患GL

「非劣性試験では、重大な試験逸脱による影響を検討すべき」

論点3：非劣性試験

- E9：同等性試験又は非劣性試験では、最大の解析対象集団を使用することは一般に保守的ではないため、その役割は十分慎重に考慮すべきである。

PPSを用いるべきという非統計家の誤解

- E9：最大の解析対象集団の解析と治験実施計画書に適合した対象集団の解析が本質的に同じ結論に達する場合、試験結果の信用度は高くなる。しかしその場合でも、最大の解析対象集団の相当の割合を治験実施計画書に適合した解析で除かなければならない場合には、試験全体の妥当性が疑われることに注意する必要がある

安定性が増すとは書かれていない

論点3 補足資料(主要層別とは)

- 中間事象が発現する可能性に応じて、次の4つの部分集団を定義(興味のある集団は S_{11})
 - S_{00} : いずれの群でも逸脱を起こすであろう集団
 - S_{01} : 新薬群では逸脱なし, 対照群ではありであろう集団
 - S_{10} : 新薬群では逸脱あり, 対照群ではなしであろう集団
 - S_{11} : いずれの群でも逸脱を起こさないであろう集団

新薬群

		新薬群	
		逸脱あり	逸脱なし
対照群	逸脱あり	S_{00}	S_{01}
	逸脱なし	S_{10}	S_{11}

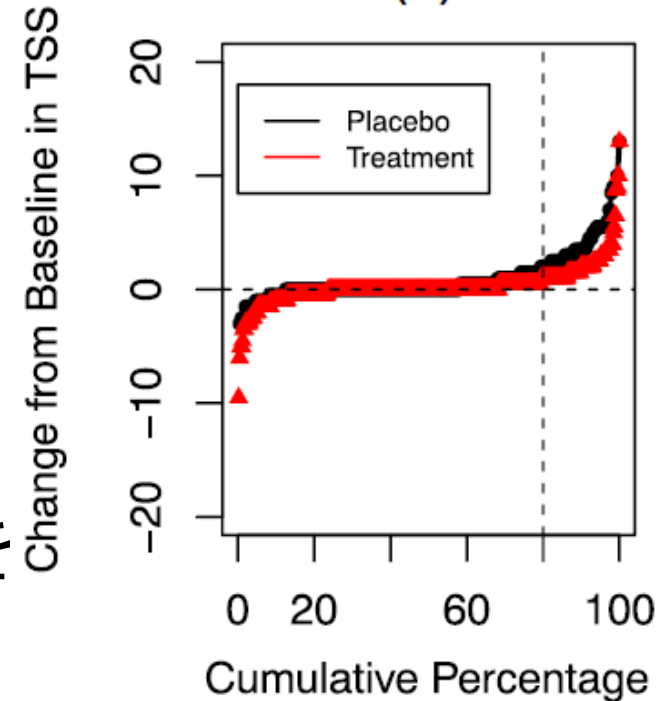
論点4 関節リウマチにおける関節の構造的損傷の進行抑制効果の評価

- 治療の進歩に伴い、関節リウマチの治療目標は疼痛除去だけではなく、関節破壊進行の阻止も求められるようになった
- Van der Heijde's modified total Sharp score (mTSS)
 - 単純X線画像所見を数値化した指標
 - 総スコアは448点
 - 値が大きいほど状態が悪いことを示す
 - 臨床試験では多くの場合、24週後のベースラインからの変化量を評価



論点4 mTSSによる評価の課題

- mTSSは24週間程度では大きく変化しない
- 各群の平均値や中央値は0
- mTSSの解析には順位に変換した解析が実施されることが多い
(Wilcoxonの順位和検定等)
- 結果として、一部の関節破壊が進行した患者での差異を評価
- 臨床的にはmTSSの群間の違いを定量的にどう解釈するかが課題



Dong and Lu. Pham Stat 2019

論点4 Dong and Luの提案と議題

- Dong and Lu (Pharm Stat 2019) は関節破壊が進行する集団を主要層として、当該集団に対する平均の差を評価することを提案

このようなケースを主要層アプローチと言って良いのか？

関節破壊の進行を中間事象としてよいのか？

論点5：試験中止は中間事象か？

- 中間事象（ICH-E9 R1 step2 用語集）
 - 治療開始後に発現し、変数を観測できなくする、又は変数の解釈に影響を与える事象。
- 欠測データ

規定したestimandの解析に対して意味があると考えられるが、収集されなかったデータ。存在しないデータや中間事象の発現により意味があると思わせないデータとは区別される。

論点5：試験中止は中間事象か？

- ICH E9(R1) Step 2 Training Material Module 1.

Summary

“Study discontinuation, loss-to-follow-up or other missing data are **not intercurrent events** and are not reflected in the estimand, but instead represent limitations to the data to be addressed in the analysis.”

試験中止は、中間事象の定義にあてはまるように見えるが、中間事象ではないのか？

事前のコメント:

#	回答(意識を含む)
1	検証試験以外でもestimandを構成すべきか
2	Estimandのプロトコールでの記載方法
3	疾患領域特有の中間事象例
4	Per protocol setの位置づけ
5	試験治療中止後のデータはtreatment strategy以外では不要か
6	Estimandの検討で問題になったこと
7	estimandを構成していない試験の事後的解釈上の問題点は？
8	仮想ストラテジーが適切な場合は？
9	主要層ストラテジーの使い方
10	開発の相とestimandの関係
11	併用禁止薬とestimandの関係
12	Estimandの非劣性試験での考え方
13	試験収支は中間事象か？