

東京大学大学院情報学環 生物統計情報学コース主催 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?

13FEB2025

ディスカッションに向けての論点整理

東京大学大学院医学系研究科
公共健康医学専攻 生物統計学
松山 裕

1

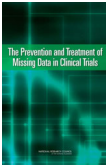
ICH E9(R1)によって日本の臨床試験の計画と解析は変わるのか?

- 1998年 (ICH) 以前
 - 1983年：新薬の臨床試験の実施に関する専門家会議の設置 (厚生省)
 - 1990年：医薬品の臨床試験の実施に関する基準 (旧GCP) 施行
 - 1991年：第1回ICH会合 (ブリュッセル)
 - 1992年：臨床試験の統計解析に関するガイドライン (旧統計ガイドライン)
- 1998年 (ICH施行) 以降
 - 1998年：医薬品の臨床試験の実施の基準に関する省令 (新GCP : ICH E6)
 - ： ICH E9 臨床試験のための統計的原則
 - 2019年：ICH E9(R1)
 - Estimandの時代？

2

2

臨床試験の計画と解析に関する基本的な考え方

- ICH E9, 1998 : Statistical Principles for Clinical Trials (臨床試験のための統計的原則)
 - ランダム化、盲検化、ITT原則、対照群の選択、多重性の調整、中間解析、...
- International Council for Harmonization (1998). "E9: Statistical Principles for Clinical Trials," available at http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf, or <https://www.pmda.go.jp/int-activities/int-harmony/ich/0031.html>
- NRC (National Research Council) Report, 2010
 - U.S. National Academy of Sciences
 - Panel on Handling Missing Data in Clinical Trials
 - Little (Chair), Cohen, D'Agostino, Dickersin, Emerson, Farrar, Frangakis, Hogan, Molenberghs, Murphy, Neaton, Retnitzky, Scharfstein, Shih, Siegel, Stern
 - The Prevention and Treatment of Missing Data in Clinical Trials
 - National Academies Press, 2010, Washington DC.
<http://www.nap.edu/catalog/12955/the-prevention-and-treatment-of-missing-data-in-clinical-trials>
- 
- ICH E9(R1), 2019 (Step 4)
 - ICH Expert working group : 2014年に開始

3

3

今でも有用だが、治療効果に関して不明瞭な概念

- Pragmatic trial (実践的試験) or Explanatory trial (説明的試験)
- Confirmatory trial (検証的試験)
 - Effectiveness (有用性) or Efficacy (有効性)
- Analysis Population (解析集団)
 - ITT集団、FAS (Full Analysis Set)、PPS (Per Protocol Set)
- PICO (PECO)
 - Population, Intervention (Exposure), Comparator, Outcome
- ...

4

4

ICH E9(R1)の主な内容

- 治療効果が意味するもの：Estimand
 - 推定値（結果）の解釈の前にestimandの設定
- 中間事象（Intercurrent events：IcEs）
- Estimandに整合した試験デザインと統計解析方法
- 結果に関する感度解析

臨床試験と日常診療では、対象者/施設選択・患者選択/除外基準・計画されたデータ測定/管理・プラセボの使用など環境や状況が大きく異なるが...
日常診療での治療を念頭においた「治療効果の推定」

5

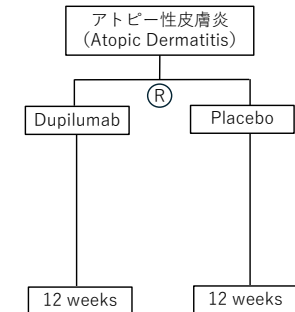
5

仮想例：ADに対する生物学的製剤（デュピルマブ）

AD (Atopic Dermatitis) アトピー性皮膚炎

- 目的
 - 新規の生物学的製剤の12週間投与とプラセボの比較
- 主要評価項目
 - 12週目でのEASIスコア

EASI Eczema Area and Severity Index (湿疹面積・重症度指数)
頭部、上肢、下肢、体幹の4つの部位の合計点（高いほど重症）



6

6

これまでの解析では...

- 論文でよく見かける文章
 - Data were analyzed on an intention-to-treat basis. Outcome data after receipt of rescue treatment or discontinuation of dupilumab was treated as missing, and a mixed model for repeated measures was used to estimate the treatment effect.
- ITT原則に従うというなら...
 - レスキュー治療開始後、あるいは治療中止後のアウトカムデータをすべて測定し、混合モデルによるデータ解析を行わなければならない
- 欠測値を含んだデータに混合モデルを用いるというのなら...
 - ITT治療効果ではない
 - レスキュー治療や治療中止がなかった場合の治療効果を求めていることになる
- いったいどんな治療効果を求めたいのか？

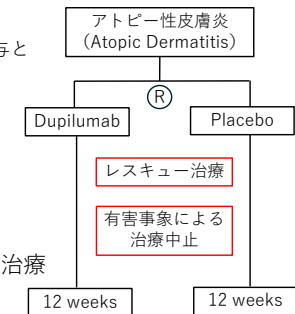
7

7

仮想例：ADに対する生物学的製剤（デュピルマブ）

- 目的
 - 「日常診療」における新規の生物学的製剤の12週間投与とプラセボの比較
- 主要評価項目
 - 12週目でのEASIスコア

EASI Eczema Area and Severity Index (湿疹面積・重症度指数)
頭部、上肢、下肢、体幹の4つの部位の合計点（高いほど重症）
- 割付治療で効果がみられない場合にはレスキュー治療
 - ただし、日常診療でもその治療は使用できるとする



8

8

どんな治療効果 (Estimand) を求めたいのか?

- これまでの (あいまいな) 治療効果の定義
 - 12週目時点での新規生物学的製剤とプラセボの改善度の違い
 - レスキュー治療や治療中止といった「**中間事象 (intercurrent events)**」が結果の解釈に影響を与えていることを考慮した記述ではない
- 例えば、以下のように詳細に記述
 - レスキュー治療の使用、あるいは治療の中止に関わらず、12週目時点での改善度の違い
 - 中間事象以降の治療状態は任意の場合の治療効果
 - レスキュー治療の使用、あるいは治療の中止がなかった場合に観察される12週目時点での改善度の違い
 - 有害事象発現がない限り、全員が予定通りの治療を受け続けた時に観察される12週目時点での改善度の違い (レスキュー治療がなかった場合の効果)
 - ...

9

9

Estimand → Estimator → Estimate



10

10

Estimandの構築

Estimand A precise description of the treatment effect reflecting the clinical question posed by the trial objective. It summarises at a population-level what the outcomes would be in the same patients under different treatment conditions being compared.

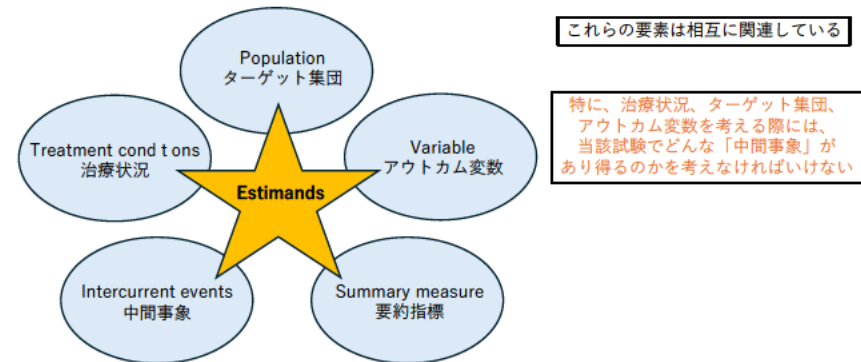
Estimand definition, ICH E9(R1) addendum

- 様々な要素について考える必要がある
- 例えば、
 - 対象疾患
 - 臨床的な状況
 - 例えば、代替治療が存在するのか/しないのか?
 - 治療方法
 - 例えば、1回きり/継続投与なのか?
 - 治療の目的
 - 例えば、予防/症状コントロール/イベント抑制なのか?
 - 試験の目的
 - 例えば、優越性/非劣性?
 - ...

11

11

Estimandにおける5つの要素



12

12

中間事象 (Intercurrent Events : IcEs)

- E9(R1) addendum
 - Events occurring after treatment initiation that affect either the interpretation or the existence of the measurements associated with the clinical question of interest. Intercurrent events should be addressed when describing the clinical question of interest to precisely define the treatment effect that is to be estimated.
→ アウトカムデータの解釈や存在に影響を与える「ランダム化後のイベント」
- 大きく分けて2種類
 - 治療内容を変化させるイベント (**treatment-modifying events**)
 - レスキュー治療の開始、副作用による治療中止、治療のスイッチング、予定外の用量の投与、...
 - アウトカムを観測不能にするイベント (**truncating events**)
 - 死亡 (例えば、エンドステージがん患者におけるQOL評価、四肢の症状スコア評価における切断、出生体重の評価における流産、...)

13

13

中間事象 (IcEs) とプロトコル違反・欠測データ

- プロトコル違反 Protocol deviations
 - 全てのプロトコル違反が中間事象になるわけではない
 - 当該のプロトコル違反が割付治療の変更に影響を与えるのであればIcEsだが (例えば、プロトコルで禁止されている治療法の使用)、そうでないならIcEsではない (例えば、同意未取得)
 - 全ての中間事象がプロトコル違反になるわけではない
 - 当該の中間事象がプロトコルで許容されていれば違反ではないが (例えば、プロトコルで規定されたレスキュー治療の使用)、プロトコルで規定されていない治療法の使用は違反
- 欠測データ Missing data
 - ターゲットestimandを推定するために意味のあるデータだが、収集されなかった
 - 追跡不能 (loss to follow-up)、試験からの脱落 (study withdrawal)、いわゆる欠測値はIcEsでは「ない」
 - IcEsと欠測データを区別するために、欠測理由を収集しなければいけない
 - 例えば、追跡不能 (単なる脱落? 効果がみられないために治療中止して脱落?)
 - Estimandの構成とは無関係であり、欠測データの対処方法は別途検討

14

14

ターゲット集団

- 治療効果を知りたい患者集団
 - 実際に試験に登録された患者だけでなく、より広い集団
 - 解析対象集団のことではない
- 研究計画書の参加適格/除外基準で明示
 - ベースライン患者特性でさらに規定されることもある
 - 例えば、サブグループ集団
 - 潜在的な中間事象で定義されることもある (主要層別 : principal stratification)
 - 例えば、どちらの治療を受けても割付治療を守る集団



15

15

アウトカム変数

- 各患者に対して測定されるエンドポイント
 - 臨床上もっとも重要な評価項目
- 中間事象の取り扱いによって変化することがある
 - 複合エンドポイントストラテジー (Composite Strategy)
 - 例えば、レスキュー治療を受けた患者は「悪化」、「最悪値」とする
 - 治療下ストラテジー (While-on-treatment Strategy)
 - 例えば、レスキュー治療を受けた患者はその直前までのアウトカムとする



16

16

治療状況



- 当該試験で比較したい治療方針（戦略、レジメン）
- 例えば、
 - 新治療A vs 標準治療B
 - 新治療A + レスキュー治療 vs 標準治療B + レスキュー治療
- どのような中間事象が考えられ、その中間事象にどのように対処するのか
 - 中間事象が日常診療では避けられるのか
 - 対処方法が日常診療でも可能かどうか
 - 対処方法は当該疾患に対してどの程度影響するのか

17

17

例えば、ADに対する生物学的製剤の仮想例で...

- 目的
 - 「日常診療」における新規の生物学的製剤の12週間投与とプラセボの比較
- 中間事象1 (IcE1)
 - 日常診療でも利用可能なプロトコルで規定されたレスキュー治療の使用
- 中間事象2 (IcE2)
 - デュピルマブに起因した有害事象による治療中止
- 治療方針
 - レスキュー治療の使用、あるいは有害事象による治療の中止に関わらず、新規生物学的製剤とプラセボの12週間投与

18

18

Estimandにおける5つの要素

要素	内容の記述
治療状況	レスキュー治療の使用、あるいは有害事象による治療の中止に関わらず、新規生物学的製剤とプラセボの12週間投与
ターゲット集団	参加適格基準に合致する確定診断を受けたアトピー性皮膚炎患者
アウトカム変数	EASIスコアで測定された重症度の変化（12週時点-ランダム化時点）
要約指標	比較群間でのアウトカム変数の平均値の差

- IcEsの取り扱い、治療方針ストラテジー（Treatment Policy Strategy）
- したがって、アウトカム変数はIcEsとは関係なく全員12週目まで測定する必要がある

19

19

例えば、IcE3：外用療法（塗り薬）の使用

- 中間事象に対する5つのストラテジー
 - Strategies for Addressing Intercurrent Events when Defining the Clinical Question of Interest (A.3.2)
1. 治療方針（Treatment policy）
 2. 仮想（Hypothetical）
 3. 複合変数（Composite）
 4. 治療下（While-on-treatment/alive）
 5. 主要層（Principal stratum）

5つが併記されているが...
当該試験でどれか1つを採用する
というものではない
中間事象の内容によって採用する
ものは変わる
アウトカムの種類によっても変わる

20

20

治療方針ストラテジー

- 中間事象の発生に関わらず、その影響も加味した治療効果
 - 中間事象は治療の一部と捉える
- いわゆるITT治療効果
 - 中間事象発生以降のデータも測定する必要がある
- Treatment-modifying eventsには汎用性があるが、治療効果の意味は？
 - 中間事象の発現状況が当該試験と日常診療で同じだった場合に観察される効果
 - 本来は、ICEsへの対処方法がプロトコルで規定されている必要がある
 - オプション1：塗り薬の使用は任意
- Truncating eventsには対応できない
 - 例えば、死亡後のスコアは存在しない

21

21

仮想ストラテジー

- 中間事象の発生がなかった場合に観察される治療効果
 - いわゆる有効性 (efficacy) の評価
 - 中間事象の発現以降のアウトカムを明示/暗示的に予測あるいは補完する必要がある
 - 中間事象発生以降のデータの測定は必要な時もあるれば不必要な時もある (推定方法に依存)
- Treatment-modifying events/Truncating events
 - 合理的なシナリオなのか、臨床的に重要なシナリオなのかを正当化する必要がある
 - レスキュー治療がなかった場合
 - 有害事象がなかった場合
 - 死亡がなかった場合
 - ...
 - オプション2：塗り薬の使用がなかった場合の治療効果

22

22

複合変数ストラテジー

- 中間事象の発生はアウトカム変数に含まれる
 - 中間事象を発生した対象者のアウトカムには特定の値を与える
 - 例えば、中間事象を発生した対象者は、無効 (non-responder)、最悪値など
 - 中間事象発生以降のデータは測定する必要はない
- Treatment-modifying events/Truncating events
 - いずれにも対応可能だが...
 - 結果の解釈が困難になる場合がある
 - 平均値にはバイアスがある
 - アウトカム変数の構成要素間で治療効果が異なる (逆転した) 場合
 - 一方の治療群のみで発生する中間事象の場合はアウトカムの定義が群間でそもそも異なる
 - オプション3：塗り薬を使用すればEASIスコアを加点 (あるいは、最悪値)

23

23

治療下ストラテジー

- 中間事象の発生より前に測定されたアウトカムデータを使用
 - 治療を受けている (あるいは、生存している) 間の治療効果
 - 中間事象発生以降のデータは測定する必要はない
- Treatment-modifying events/Truncating events
 - いずれにも対応可能だが...
 - 治療期間が重要でない、あるいはイベント発生率が一定のときには適切かもしれない
 - 結果の解釈が困難になる場合がある
 - 治療群間でのアウトカム評価時点が異なる (いわゆるLOCF) LOCF: last observation carried forward
 - 中間事象の発生が比較群間で異なる場合
 - 治療群間で中間事象発生状況がどの様に違うのかを考慮して結果を解釈する必要あり
 - オプション4：塗り薬を使用する直前のEASIスコアを使用

24

24

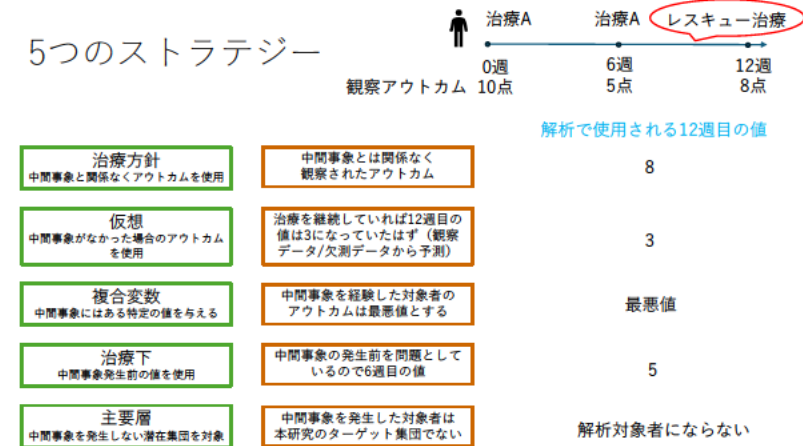
主要層ストラテジー

- 関心のある主要層における治療効果
 - 実際に受けた治療に関わらず、中間事象の発生状況が同じ「潜在集団」
 - 中間事象発生以降のデータは測定する必要はない
- Treatment-modifying events/Truncating events
 - いずれにも対応可能で、そのような状況に関心がある場合もあるが...
 - 抗がん剤に反応した（腫瘍縮小が観察された）集団における奏効期間の比較
 - エンドステージがん患者におけるQOL評価、...
 - そもそもそのような潜在集団は同定不能
 - 仮想的なターゲット集団
 - オプション5：どちらの治療法に割り付けられても塗り薬を使用しない集団で解析

25

25

5つのストラテジー



26

26

ADに対する生物学的製剤の仮想例では...

- 想定される中間事象とそれらへの対応
 - レスキュー治療の使用：lcE1
 - レスキュー治療が日常診療でも使用可能か/その影響はどの程度なのか？
 - デュピルマブに起因した有害事象による治療中止：lcE2
 - 日常診療で有害事象を軽減しながらデュピルマブ投与を継続可能かどうか？
 - 外用療法（塗り薬）の使用：lcE3
 - 日常的に使用するかどうか/その影響はどの程度か？
- 例えば、以下の状況であれば...
 - プラセボ対照の薬効評価を治験で行っている
 - 有害事象を軽減できる方法はない
 - 塗り薬は日常的に使用し、病状に対する影響は小さい

27

27

Estimandにおける5つの要素

要素	内容の記述	
治療状況	有害事象発現と塗り薬の使用に関わらず、レスキュー治療がなかった場合の新規生物学的製剤とプラセボの12週間投与	lcE1, lcE2, lcE3
ターゲット集団	参加適格基準に合致する確定診断を受けたアトピー性皮膚炎患者	
アウトカム変数	EASIスコアで測定された重症度の変化（12週時点-ランダム化時点）	
要約指標	比較群間でのアウトカム変数の平均値の差	

- lcE1：仮想ストラテジー
- lcE2とlcE3：治療方針ストラテジー
- アウトカム変数はlcE2とlcE3と関係なく12週目まで測定し、解析に使用する必要がある

28

28

中間事象1：Treatment-modifying events

- 当該試験においてどのようなものがあり得るか/それらへの対応は?
 - 実際には様々な種類や予想できないものも考えられるが、程度問題だと思う...
- 大きくは3種類
 - レスキュー治療や後治療（治療スイッチングを含む）の開始
 - 開始のための明確な基準設定
 - 実臨床でも実施可能な合意された治療方針なのか?
 - 有害事象による治療中止
 - 軽減のための治療（対処）法が存在するか/しないか
 - 併用禁止薬の使用
 - 作用機序の異なる他剤や当該薬剤の効果を増強させるような薬剤の使用
 - レスキュー治療に含めて考えればよい
 - その他、有害事象を引き起こす可能性がある薬剤の使用など
 - 試験の質管理で減らすことができるはず

29

29

中間事象2：Truncating events

- 当該試験においてどのようなものがあり得るか/それらへの対応は?
 - 実際には色々あると考えられるが、程度問題だと思う...
 - よく出てくる例は、symptom-related trials*における死亡イベント
 - * 症状改善を目的とした代替エンドポイントに対する試験
- Truncating eventsが多数観察される試験
 - エンドポイントの一つではないか（複合変数ストラテジー）
 - Outcome trials*ではこれまでも複合エンドポイントにしている試験が多い
 - * 真のエンドポイントに対する大規模長期イベント試験
 - 四肢の症状スコア評価における四肢切断：最悪値
 - エンドステージがん患者におけるQOL評価の際の死亡
 - 主要層ストラテジー? 複合変数ストラテジー?
 - 死亡イベントが多い状況であれば、生存時間とQOLの両方をアウトカムと考えるべきでは?
 - 競合リスクの問題
 - エンドポイントの観察を不能にするイベント

30

30

試験の計画と実施に対する影響

- 試験の計画
 - 試験の目的を反映したestimandと整合性を保つ必要がある
- プロトコルで記載すべき試験デザイン（ICH E6）
 - 試験のタイプ（例：二重盲検、プラセボ対照、並行群間、優越性/非劣性、...）
 - 対象者の観察期間
 - 治療中止の基準、試験中止の基準
 - 試験期間前・中に許容される治療法（レスキュー治療）
 - 治療遵守状況のモニタリング方法
 - 有効性/安全性に関わる臨床パラメータの評価/記録の時期/方法
- サンプルサイズ設計
 - ターゲットのestimandに対する適切な効果量とバラツキを採用しているか?

Cook JA, et al, 2018

31

31

用語について

- 5つのストラテジーの用語にこだわる必要はないが...
- 治療方針 = ITT } ではない
- 仮想 = 因果効果 }
- 因果推論の枠組みのもとでは、全てのestimandは治療方針の比較であり、仮想的な治療効果
 - Efficacy（有効性）とEffectiveness（有用性）のどちらも統一的に表現可能
 - Per-protocol causal estimand

Hernán, Scharfstein, 2018

32

32

Per-protocol causal estimand

- 因果治療効果 (Causal estimands/treatment effects)
 - 明確に定義された患者集団において実際に実行可能な治療戦略の対比 (比較)
 - Per-protocol causal estimands: $E(Y_{g=a1}) - E(Y_{g=a0})$
 - $Y_{g=a}$: $g=a$ という治療方針を受けたときの反事実アウトカム
 - 仮想的な治療効果
 - 識別可能にするためにはデータから検証不能な仮定が必要
 - ただし、対象集団は現実のもの (測定されたベースライン特性で定義可能)
- 治療方針?
 - プロトコルで事前に明確に規定
 - 規定できれば、いわゆる中間事象の取り扱いが定まる
 - 許認可の場面では難しい (治療方針を認可するわけではない)

33

33

治療方針の例

- 中間事象の内容によらず、ずっと割付治療を投与
 - なんとなく「因果効果」と思われている戦略
 - 全ての中間事象がなかった場合に観察される治療効果
- 中間事象発現以降は任意の治療法
 - ITT治療効果
- 有害事象による治療中止がない限り、割付治療を投与
 - 有害事象は実臨床でも避けられないのであれば、「有害事象がなかった場合」を考えることには意味がない
- 動的治療レジメン (dynamic treatment regime)
 - 例えば、薬剤XとPの比較で状態悪化後は治療B (B1, B2) のいずれか
 - Treat X until progress on, then treat B2 vs. Treat P until progress on, then treat B2
- ...

34

34

Outcome trialsとSymptom-related trials

- 真のエンドポイントに対する大規模長期イベント試験
 - 中間事象の発生は不可避
 - 複合変数と治療方針ストラテジーの組み合わせで対応するのが標準的
 - しかし、そのような治療効果に意味があるのか?
 - 複合イベントの解釈
 - Black-boxな治療方針であるITT治療効果
- 症状改善を目的とした代替エンドポイントに対する試験
 - 中間事象は発生するけれども...
 - プラセボを使用する場合であれば、薬効評価が第一義的では?
 - 仮想・治療方針・複合ストラテジーの組み合わせ

35

35

治療方針ストラテジー

- ITT検定
 - ランダムでない中間事象があったとしても、妥当な α レベル検定
- ITT治療効果
 - 治療方針の比較?
 - 治療方針: "Assign to treatment, then do whatever you want"
 - "Black-box" 治療方針なので、因果効果としては不十分
 - 正確には、臨床試験で観察された治療法変更内容・割合、治療法変更に影響を与える要因などが実臨床でも同じだと仮定した場合に観察される効果
 - 仮想的な治療効果
 - なんでもかんでもITT治療効果では問題あり

36

36

仮想ストラテジー

- ICH E9(R1)では、否定的な文章も見受けられるが...
 - プラセボの使用を含む日常診療とは乖離した状況での薬効評価を考えれば第一義的なestimandではないか?
 - 中間事象がなかった場合の治療効果
 - ただし、中間事象がなかった場合というのが医学的に合理的なシナリオなのか?
 - 例えば、
 - レスキュー治療がなかった場合の治療効果
 - 有害事象による治療中止がなかった場合の治療効果
 - 全ての対象者が治療を完遂できた場合の治療効果
 - ...

37

37

感度分析

Sensitivity analysis: A series of analyses conducted with the intent to explore the robustness of inferences from the main estimator to deviations from its underlying modelling assumptions and limitations in the data.

Sensitivity analysis definition, ICH E9(R1) addendum

- メインの推定量に対する事前に計画された感度分析
 - 特定の推定方法における仮定に対する結果の頑健性の検討
 - 推定方法の選択に対する感度分析ではない (例えば、MARのもとでMMRM vs. MI)
 - どの程度の仮定からの乖離であれば、結果の解釈 (統計的・臨床的有意性) を変化させるのか?
 - 例えば、欠測値に対するMARの仮定
MNARのもとでの治療効果の変化 (Tipping point analysis)
- Estimandの種類によって重要性はことなる
 - 治療方針：データ収集をきっちり行うことで感度分析の役割は少ない
 - 複合変数・治療下：統計的な仮定は少ないかもしれないが、解釈上の感度分析
 - 仮想・主要層：データからは検証不能な仮定に依存

38

38

試験デザインとデータ収集

- 試験デザイン
 - どのような中間事象がありえるのか?
 - 予想される中間事象への医学的対応方法は?
 - その対応方法がアウトカムに与える影響はどの程度なのか?
 - 対象者/施設選択・患者選択/除外基準など臨床試験と日常診療との乖離は?
 - ...
- データ収集
 - 欠測値と区別するためにも中間事象が生じた詳細な理由を収集する必要がある
 - Estimandの種類によって中間事象以降のデータ収集の必要性・重要性はかわる
 - ...

39

39

Estimandの議論

- 臨床試験の計画・実施・解析・報告に関与する臨床家、統計家およびその他の専門分野の担当者も含む多分野にわたる作業が必要
 - 非常に喜ばしいこと
- 臨床試験の様々な側面からの議論が必要
 - 統計的側面：当然
 - 臨床的側面：試験の性質、疾患の特性、治療体系、中間事象の種類、...
- まずは、共通認識 (言語) をもつことが大事
 - 議論が成立しないことがある

40

40

ご清聴ありがとうございました

41