

当日のディスカッション

【テーマ4】エンドポイント間の予測と
代替性の評価

事例1: エンドポイント間の予測 (Ph2の結果からPh3の治療効果を見極めるには?)

• ディスカッションポイント

- Ph2の結果からPh3の治療効果を見極める方法(今回はベイズ流メタ回帰)は有用か?
- 治療効果を固定した検出力に対して, 治療効果のばらつきを踏まえて検出力の期待値をとったPoS (Probability of Success)* という考え方は有用か?
 - *なお, successは有意/非有意以外にも設定可能であるため, 検出力の期待値をとったものだけがPoSというわけではない

• メタアナリシス(メタ回帰)に関して...

- 収集する基準の必要性(治療ライン, 標準治療を統一する/しない, 年代, 人種)
 - 標準治療は統一することが多いが, それがベストかの答えはない
 - 臨床的な背景も含めてデータを集めないと説明ができなくなる
- そもそも対象となるエビデンスが少ない場合も(分子標的薬/対象集団の制限)
- 事例の結果は綺麗すぎるのでは?
- 明確な関係性が期待できないが, 真のエンドポイントの実施可能性が低い場合は, サロゲイトでも十分評価可能とする場合もあるのでは?
(例: 一部のがん種におけるPFSとOS)
- データ収集が難しい場合, 収集済みの外部データを活用することもある
 - Berry et al. (2017) Creating a synthetic control arm from previous clinical trials: Application to establishing early end points as indicators of overall survival in acute myeloid leukemia
- 治療体系の変化が激しい領域では, 利用しづらいかも

事例1: エンドポイント間の予測 (Ph2の結果からPh3の治療効果を見極めるには?)

- 手法の仮定に関する課題
 - ベイズにおける事前分布の仮定
 - 複数の事前分布による感度分析(効果がないとする分布, 意義ある値を持つ分布)
 - ハザード比を利用する場合の各試験の比例ハザード性
 - RMSTなど他の指標を利用すれば問題とはならない
 - 論文に記載がなかったとしてもKMプロットから生データを再現, 検討する場合も
- Probability of successに関する意見
 - 振り返りとして取り入れている企業もある
 - 算出された値の判断基準に困る場合がある
 - シミュレーションによる性能評価を紹介
 - 【After meeting memo】データサイエンスラウンドテーブル第4回 テーマ4 ベイズも参考となる
 - <https://www.pmda.go.jp/files/000217614.pdf>
 - <https://www.pmda.go.jp/files/000217616.pdf>
- 予測だけでOKか? 代替性も必要か? という意見
- 事例: 喉の気圧差と呼吸能力を患者データでモデル化。健康成人から得られたデータ(喉の気圧差)で, 患者における治療効果(呼吸能力)を予測
 - 必ずしもすべての疾患に該当するわけではない

事例2: エンドポイントの代替性評価(単一試験)

- 代替性とは？

- 「因果関係ベース(Z (投与群) $\rightarrow S \rightarrow T$ かつ $Z \rightarrow T$ はなし)」か「 S の差で T の差を予測できるかどうか？」
 - サロゲイトエンドポイントとバイオマーカーの違い
(バイオマーカーの場合は必ずしも因果関係は不要だが、サロゲイトは因果関係が必要という認識だった)
- 臨床的なメカニズムが説明できるのであれば、データに基づく因果関係の評価は不要で、関連だけでも十分では？
 - データに基づく因果関係の評価は重要だが難しい
- 実際は様々な評価項目で(主に関連について)網羅的に検討されていることが多い
- 臨床家にとって何を評価しているかが理解しやすいことも重要

- 各指標に関して・・・

- Prenticeの4基準は何が厳しいのか？
 - 検定ベースだと4つ目の基準(投与群から真のエンドポイントへの直接効果がない)の検証が難しい
 - だったらベイズなら？(Bayes factorの利用)
- PTEは何が厳しいのか？
 - 「PTEの95%CIの下限が0.5を超えることが望ましい」という基準
- REの結果は外挿することが難しい
- Trial-levelとindividual-levelの乖離
 - Individual-levelの関連がなくても、trial-levelの関連があればよいのでは？
 - 臨床的なメカニズムがあれば、上記のように個人の関連がないことはあまり考えられない
 - Individual-levelの関連だけがない場合は、データ数の影響が考えられる

事例2: エンドポイントの代替性評価(単一試験)

- そもそもなぜサロゲイトエンドポイントが必要なのか？
 - 知れるのであれば早く知りたいという動機
- サロゲイトエンドポイントの決め方
 - 複数の介入試験から決めるべきか？単一試験でもOKか？試験数は？
 - システムティックレビューはした方が良い
 - エビデンスとして広範囲の方が一般化可能性がある
 - 一方で実施可能性も考慮すべき
 - カットオフの決め方
 - 事例3のやりかたが参考になるかも

事例3: エンドポイントの代替性評価(複数試験)

- ディスカッションポイント
 - 事例として示した結果が提示された場合、検討内容及び検討結果が充足していると考えるか否か？
- データセットの適切性
 - 古い年代が入っていて、年代ごとの治療の変化が考慮されていないので、年次効果を考慮した方が良いかもしれない
 - Surrogacyの精度はfirst lineとlast lineで分けても良いのではないか？
 - ステージ2と3の代替性の結果の違いの原因を説明する必要がある
- 部分集団をどこまで検討すべきか？(事例のステージ別以外)
 - 代替性の結果を使う目的による。その集団での開発を目指しており、代替エンドポイントを用いたいのであれば、代替性評価をしておくべき
 - 個人レベルで見たいのか？集団レベルで見たいのか？も重要
 - 個人レベルならデータベースから似たような人をマッチングするだけで良いかも

事例3: エンドポイントの代替性評価(複数試験)

- 代替性評価をするにあたって・・・
 - フォローアップ期間の設定方法
 - 比較的似ている他の試験を参考にする
 - 測定ポイント
 - Surrogacyのために介入するのではなく, 臨床的に実施可能な範囲でデータを取得した上で, 利用可能なデータでsurrogacyを検討する
- Subject-levelのデータの利用可能性
 - 製薬協レポートがまとめられた頃(2009年)と比較して, subject-levelのデータがより利用可能となってきている
 - 例: Clinical Study Data Request
<https://clinicalstudydatarequest.com/>
 - 誰が何のために利用するのか? が今後重要となってくる
 - 誰が: 製薬企業, 規制当局など
 - 何のために: 「よい薬をより早く」のための, より適切なサロゲイトエンドポイントの開発, サロゲイトエンドポイントの妥当性の評価