

第9回データサイエンスラウンドテーブル会議

仮想事例についての最適な試験デザインの計画

野村 尚吾（東京大学）

中村 将俊（ファイザー）

澤本 涼（中外製薬）

萩原 駿祐（小野薬品）

青木 誠（ノバルティスファーマ）

浅野 淳一（PMDA）

佐々木 誠治（PMDA）

仮想事例1

仮想事例1の背景 (1)

- 神経芽腫の高リスク群に対して、再発を防ぐための地固め療法後の追加の化学療法として、新たな薬剤を開発する想定
 - 神経芽腫は小児がんの1つで、日本の年間患者発生数は150人/年と想定
 - そのうち高リスク群で地固め療法まで実施する患者は、20～30人/年と想定
- 同じ対象に対して、外部対照群との比較を有効性の主たる評価とした事例が2つ存在
 - Eflornithine (IWILFIN)
 - 外部対照データはANBL0032試験 (アメリカで2009年から2015年に組み入れられた患者を10年間追跡)
 - ¹³¹I-Omburtamab
 - 外部対照データはCentral German Childhood Cancer (CGCC) レジストリ (ドイツで1990年から2015年に診断された患者が登録)

仮想事例1の背景 (2)

- 仮想的な薬剤としてヨクナルマブという名前の新規化合物を想定
 - 他のがんや疾患で承認はされていない。
 - 第I相を試験を終了し、MTDは決定済み
- ランダム化比較試験 (RCT) を実施する場合の必要サンプルサイズは1群166例 (2群332例)
 - 主要評価項目は、この領域で標準的なエンドポイントであるEvent Free Survival (EFS)
 - ANBL0032試験の結果を参考に標準治療の2年時のEFS率を45%、ヨクナルマブの2年時の期待EFS率を60% (期待ハザード比を0.64) と設定したとき、有意水準両側5%、検出力80%のもとで必要サンプルサイズ
- ANBL0032試験とCGCCLレジストリは利用可能と想定

**ディスカッションでは、本剤の有効性を検証する試験を実施する上で、
案1：RCT、案2：傾向スコアを用いた外部比較、案3：ハイブリッドコントロールを用いた比較、
案4：閾値に対する単群試験の中から、どれを選択するかを議論した。**

提案する試験デザイン	<デザイン名>	
開発の相	Phase 3	
主要なEstimand	<ul style="list-style-type: none"> ・治療: ヨクナルマブvs標準治療 ・集団: ランダム化された集団 ・エンドポイント: EFS ・変数の集団レベルでの要約: ハザード比 	<div style="border: 2px solid red; padding: 10px; display: inline-block;"> <h1 style="margin: 0;">Group 1</h1> </div>
試験デザインの詳細	<ul style="list-style-type: none"> ・ハイブリッドコントロールを用いる <ul style="list-style-type: none"> ・海外の外部データを単純に利用するのは難しい(人種、時期の違い) ・日本人の標準治療のデータを含められるのが利点 ・傾向スコアでマッチングする ・解析はベイズで動的利用する <ul style="list-style-type: none"> ・Type1 errorのインフレの上限をできるだけ検討する(が、エンドポイントによっては不可能) 	
この試験デザインに必要な前提条件	得られる結果の解釈	
開発予算は潤沢にあり、早期に承認を得たいものと仮定	もし得られた結果に対する解釈が、何らかの条件・状況によって変わってくるのであれば記載してください。	

提案する試験デザイン	ハイブリッドコントロール	
開発の相	Phase 3	
主要なEstimand	<ul style="list-style-type: none"> ・EFS (増悪、再発、2次がん、死亡) ・secondでOS 	<div style="border: 2px solid red; padding: 10px; display: inline-block;"> <h2 style="margin: 0; color: red;">Group 2</h2> </div>
試験デザインの詳細	<ul style="list-style-type: none"> ・対象集団: 小児がん患者 ・対称群: 日常診療のハイブリッドコントロール ・中間解析 (安全性、early stop、対照群のサイズ) ・ランダム化 ・サンプルサイズ: ICH E1 GLから被験群で100例 	

この試験デザインに必要な前提条件	得られる結果の解釈
<ul style="list-style-type: none"> ・標準治療確立していない ⇒ 早期承認取得を目指す ・外部対照群を確認 (フィジビリティ調査) ・安全性への懸念 	<ul style="list-style-type: none"> ・early stop出来る。 ・外部対照とマッチングしていない場合、普通のRCTになる ・年度ごとの治療成績の影響を受けない

提案する試験デザイン	<デザイン名>	
開発の相	Phase 3	
主要なEstimand	対象患者: 高リスク群 主要評価項目: EFS 比較: 比例ハザードモデル 中間事象:	<div style="border: 2px solid red; padding: 10px; display: inline-block; color: red; font-weight: bold; font-size: 24px;">Group 3</div>
試験デザインの詳細	国際共同治験 ハイブリッドコントロールを用いた比較 初期割付比: 1:1 (当初の割付比を1:1にして, 中間解析で類似度を評価して割付比を変更する) 層別因子: 地域 想定サンプルサイズ: 約200	
この試験デザインに必要な前提条件	得られる結果の解釈	
試験期間中に新規治療が出ない 地域間の一貫性が保証される ドイツの事例をふまえると, 時期による差は大きい可能性を確認する必要がある 対照群と外部データの異質性があまりない ANBL試験はアメリカの患者だけなので, ハイブリッドにすることで, RCT単体の結果を見る	RCTの国外のデータと外部データの一貫性が保たれないと解釈が困難になる 前提条件が成立していればサンプルサイズを抑えながら結果を得られる	

仮想事例2

仮想事例 2 の背景 (1)

- 多発性骨髄腫 (multiple myeloma : MM) の治療薬として、新たな薬剤を開発する想定。
- MM領域での薬剤開発は、従来通りのOS/PFSをエンドポイントとしたRCTを計画・デザインした場合、過度な期間や労力がかかることで、試験にかかるコストの面や患者さんに薬を届けるまでに時間がかかる倫理的な面からも、従来通りに必要なRCTを実施する事が、必ずしも最適なデザインとはいえない状況が考えられる。
- MMに対する治療効果の評価について、従来のOS, PFSによる評価の他に、Minimal Residual Disease (MRD) を用いた病態の評価が検討して貰う。
 - FDAのODACで公開された資料を基に、MRDのサロガシーについてのメタアナリシスを紹介した。
 - 複数のエンドポイントを一つの試験で評価するための多重性の調整方法や、同時に取り扱うことが可能な2in1 Adaptive designについて紹介した。

仮想事例 2 の背景 (2)

- MM治療に対する新規薬効成分であるアンゼンタマブの初回申請を想定する。
 - アンゼンタマブは、他の癌腫での承認があり、使用実績は蓄積されている。
 - 類似成分の薬剤が新たに承認されたばかり。
- アンゼンタマブの至適用量については特定され、有効性について評価する検証試験を検討。
- アンゼンタマブの薬物性由来の安全性や服薬利便性については明確に類薬より改善されている。

ディスカッションでは、類薬との比較において各Endpointに対する事前の想定は下記の状況を想定し、MRDをPrimary Endpointに設定した上で本剤の有効性を検証する試験を実施を検討した。

MRD@12mo : 群間に大きな差が見込める、検出力PFS・OSと比較してかなり高い、解析時期も早い
PFS : 群間に大きな差が見込める、検出力はOSと比較して高い、MRDの後OSより早く解析可能
OS : 群間に大きな差が見込めない、検出力はかなり低くイベントも起こりづらく解析に時間がかかる

提案する試験デザイン	<デザイン名>	
開発の相	Phase 2/3	
主要なEstimand	中間: 評価項目: MRD 早期有効中止の条件: MRD 早期中止後もPFS、OSをフォローアップ 例数: 150-200で実施 最終: 評価項目: PFS 例数: 400 - 500 (PFS想定で設計)	<div style="border: 2px solid red; padding: 10px; display: inline-block;"> <h1 style="margin: 0;">Group 1</h1> </div>
試験デザインの詳細	例数設計はPFSを前提に行い、主要評価項目をMRDとして中間解析を行う。 MRDで有効の場合、早期中止 (PFS、OSはフォローアップ) MRD: 中間Neg → 終了時MRDで評価 MRD: 中間Pos → 終了時PFSで評価	
この試験デザインに必要な前提条件	得られる結果の解釈	
PFSのみで見るとは、類薬投与例が増え、対照となるSOC群が確保しにくいので、早期にMRDを使った判断をしたい。中間解析結果が出るまでに最終解析に必要な例数が登録されている前提。		

提案する試験デザイン	階層デザイン	
開発の相	Phase 3を想定(変更可)。	
主要なEstimand	<ul style="list-style-type: none"> ・Primary: MRD@12month、PFS ・Secondary: OS 	<div style="border: 2px solid red; padding: 10px; display: inline-block;"> <h2 style="margin: 0; color: red;">Group 2</h2> </div>
試験デザインの詳細	<ul style="list-style-type: none"> ・対象患者: 多発性骨髄腫 ・中間解析: 1年の時点で設定 ・介入: Phase Iの通り 	
この試験デザインに必要な前提条件	得られる結果の解釈	
<p>★特に新たな前提条件なし</p> <ul style="list-style-type: none"> ・多発性骨髄腫治療の初回申請 ・アンゼンタマブの使用実績あり ・類似成分の薬剤が新たに承認されたばかり ⇒副作用、コンプライアンスは改善。SARは同程度 	<p>中間解析: MRDで薬効評価出来ていて、MRD-PSFの相関が見れている。⇒有効中止で申請 それ以外⇒PFSで薬効評価を行うサンプルサイズで試験継続</p> <p>OSは参考値</p>	

提案する試験デザイン	<デザイン名>	
開発の相	Phase 3	
主要なEstimand	主要評価項目 : MRD@12mo, PFS	<div style="border: 2px solid red; padding: 10px; display: inline-block;"> Group 3 </div>
試験デザインの詳細	<p>中間解析: MRD@12mo (200名12か月フォローアップ時点) PFS (200名12か月フォローアップ時点)</p> <p>最終解析: PFS</p> <p>多重性の調整: αリサイクリング (0.5%(MRD@12mo), 4.5%(PFS)), PFSの中間解析はα消費関数を用いる</p> <p>サンプルサイズ: 400-500</p>	
この試験デザインに必要な前提条件	得られる結果の解釈	
MRD@12moからPFSがどの程度であるかを見積もることが可能 (根拠資料: FDA) 中間解析時点でリクルートは完了	中間解析のMRD@12moの結果が著しく効果が見られる場合, 承認申請	