

第7回 データサイエンスラウンドテーブル会議  
テーマ: Rare Diseaseを対象とする臨床試験

2023年6月30日

事例：セルリポナーゼ アルファ（ブリニューラ<sup>®</sup>）

# 目次

---

- 希少疾患領域における医薬品開発
- 事例紹介
  - 製品及び対象疾患の特徴
  - 臨床試験及び自然歴研究
  - 規制当局からの指摘事項及び見解
- ディスカッションテーマ
  - ①開発戦略について
  - ②FDAからの指摘事項への対策
  - ③外部対照の利用可能性及びバイアスの低減

参考資料: 日本製薬工業協会 (2022). Rare diseaseの治療効果の推測法.

[https://www.jpma.or.jp/information/evaluation/results/allotment/jtrngf000000085m-att/ds\\_202212\\_rare.pdf](https://www.jpma.or.jp/information/evaluation/results/allotment/jtrngf000000085m-att/ds_202212_rare.pdf)

# 希少疾患領域における医薬品開発

---

- 医薬品開発を効率的に進めるために重要とされること
    - 疾患の自然歴に関する十分な理解
    - 疾患の病態生理, 及び薬剤の作用機序に関する十分な理解
    - 臨床試験による治療の安全性を裏付けるための毒性に関する考察
    - 評価項目の選択または開発
    - 安全性及び有効性を確立するためのエビデンス
    - 医薬品開発中における製造上の留意事項
    - 患者, 介護者, 及び支援者の開発プログラムへの参加
    - 規制当局との対話
  - FDA (2019). “Rare Diseases: Common Issues in Drug Development Guidance for Industry”  
<https://www.fda.gov/regulatory-information/search-fda-guidance-documents/rare-diseases-common-issues-drug-development-guidance-industry-0>
- 本日はここに注目して議論いたします！

## 概要

---

- 対象疾患: セロイドリポフスチン症2型 (CLN2)
- 用法・用量: 脳室内投与、300mg 1回/2週
- 希少疾病用医薬品としての指定
  - 米国(2013年4月)、欧州(2013年3月)、日本(2018年9月)
- 薬事承認
  - 米国(2017年4月)、欧州(2017年5月)、日本(2019年9月)

セルリポナーゼ アルファ(ブリニューラ®)

# 製品及び対象疾患の特徴

## • 製品

- チャイニーズハムスター卵巣細胞から産生される遺伝子組み換えヒトTPP1 \*
- TPP1欠損症として知られているセロイドリポフスチン症2型 (CLN2) 治療を目的とした酵素補充療法製剤
- 脳室内投与用デバイスの植込み手術により実施

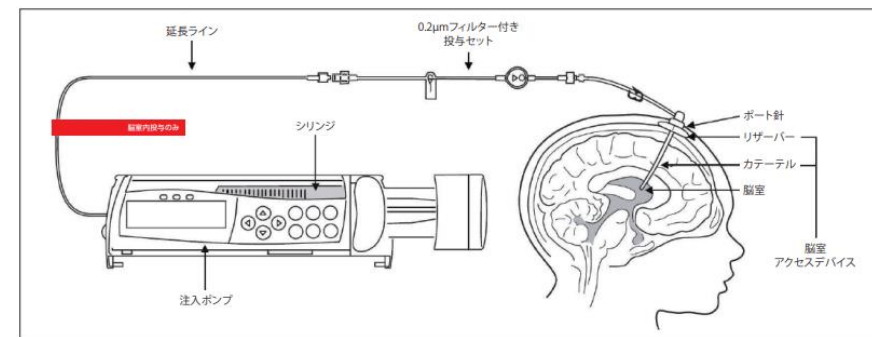
## • 対象疾患

### • セロイドリポフスチン症2型 (CLN2)

- 極めてまれな進行性の神経変性を伴う重度の遺伝子疾患
- 通常2~4歳の間に痙攣発作や運動失調、言語発達遅延を伴って発症
- 年齢を重ねるとともに多様な徴候が現れ、悪化の一途をたどる
- 末期では、失明、寝たきり、コミュニケーション不能となり多くの場合若年で死亡
- 最初の症状発現時から死亡までの期間の中央値は7.8年
- 欧米におけるCLN2の推定発症率は出生児10万人当たり0.5~1人、推定有病率は100万人当たり0.6~0.7人
- CLN2の治療法は長らく疾患進行の安定化を図るための対症療法、緩和療法のみ



投与ガイド:ブリニューラ®脳室内注射液150mg



脳室内投与システムの例

\* セリンプロテアーゼであるトリペプチジルペプチダーゼ1

# 臨床試験及び自然歴研究

---

- 有効性評価に用いられた第I/II相 臨床試験
  - 190-201試験
    - 非盲検、単群
    - 用量漸増期(4~22週間)、固定用量投与期(48週間)
    - 組み入れ患者:3~15歳のCLN2患者 24例(内日本人1例)
  - 190-202試験
    - 190-201試験の継続投与試験(最長239週)
- 自然歴研究:臨床試験での同時対照群の設定が困難であったことから実施
  - 190-901試験
    - 対象:DEM-CHILDデータベースに登録されているCLN2患者 69例
- 有効性の主要評価項目
  - CLN2評価尺度
    - 0点(重い障害)~6点(正常)
    - 運動, 言語それぞれ0~3点のスコアからなる合計スコア
  - 投与48週後に, CLN2評価尺度の運動・言語尺度がベースラインから2点以上低下しない, またはベースラインが1点の場合は0点にならない割合

セルリポナーゼ アルファ(ブリニューラ®)

## 規制当局からの指摘事項及び見解

---

- FDAからの指摘事項
  - 評価尺度の違い: 臨床試験(オリジナル尺度を改良)、自然歴研究(オリジナル尺度)
  - 評価方法の違い: 臨床試験(医師評価)、自然歴研究(レトロスペクティブな評価)
  - 評価タイミング: 臨床試験(規定時点評価)、自然歴研究(不定期評価)
  - 患者背景の違い
- 機構見解
  - CLN2運動尺度の悪化抑制は認められたが、言語尺度の有効性は確立していない
- 日本での承認審査時の対応
  - 190-901試験に新規患者を追加した集団と1対1でマッチングでの比較
- 機構見解
  - 自然歴研究との比較評価については限界があるものの、疾患の進行が抑制される傾向が示唆されている



# ディスカッションテーマ①

## • 開発戦略について

- ランダム化比較試験の実施は本当に困難か
  - 比較対照治療の設定
  - ランダム割付、盲検化
- 自然経過観察群の設定の可能性
- クロスオーバー試験の利用可能性
- 単群試験における有効性評価の妥当性
- アダプティブデザイン、ベイズ統計の適用の可能性
- その他

セルリポナーゼ アルファ(ブリニューラ®)

## 製品及び対象疾患の特徴

### • 製品

- チャイニーズハムスター卵巣細胞から産生される遺伝子組み換えヒトTPP1\*
- TPP1欠損症として知られているセロイドリポフスチン症2型(CLN2)治療を目的とした酵素補充療法製剤
- 脳室内投与用デバイスの植込み手術により実施

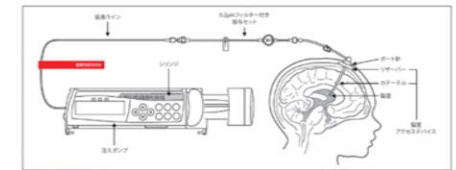


投与ガイド:ブリニューラ®脳室内注射液150mg

### • 対象疾患

#### • セロイドリポフスチン症2型(CLN2)

- 極めてまれな進行性の神経変性を伴う重度の遺伝子疾患
- 通常2~4歳の間に痙攣発作や運動失調、言語発達遅延を伴って発症
- 年齢を重ねるとともに多様な徴候が現れ、悪化の一途をたどる
- 末期では、失明、寝たきり、コミュニケーション不能となり多くの場合若年で死亡
- 最初の症状発現時から死亡までの期間の中央値は7.8年
- 欧米におけるCLN2の推定発症率は出生児10万人当たり0.5~1人、推定有病率は100万人当たり0.6~0.7人
- CLN2の治療法は長らく疾患進行の安定化を図るための対症療法、緩和療法のみ



脳室内投与システムの例

\* セリンプロテアーゼであるトリペプチジルペプチダーゼ1

## ディスカッションテーマ①

- 自然経過観察群(既存データ使用)の準備を周到にすべき。ハードエンドポイント(死亡)までは時間がかかる → 評価尺度の使用を検討する
- 自然経過観察群構築のために、データにどれだけの情報があるかを把握する
- RCTができない状況では、まず非盲検単群試験で奏功割合が臨床的に意義があるのか検証するデザインを検討する。その際、p値のみでなく推定量をベースに有効性を評価する。ベイズ統計学を用いる場合は、たとえば事前分布(無情報が望ましい)を用いて症例数を追加するごとに事後分布の信用区間の精度をあげていくデザインが考えられる。試験終了時に信用区間に基づきディスカッションする
- 単群試験でも医師が見たいものが推定できていた場合には、PMDAとも単群試験で説明できることをベースとしてディスカッションできるとよい
- 試験開始時点で治験薬の有効性に関するデータが乏しい場合は、自然経過観察群とのRCTで中間解析を実施し、治験薬の有効性が示された時点で自然経過観察群の患者にも治験薬の治療を受けてもらうデザインも検討する
- 治験薬の有効性が期待されるのであれば単群試験も許容される可能性がある。RCT(並行群間での比較)が難しい場合はクロスオーバー試験も視野に入ってくる

# セルリポナーゼ アルファ(ブリニューラ®) ディスカッションテーマ②-1

- 外部対照との比較におけるFDAからの指摘事項への対策は？
  - 評価尺度(スコア算出方法)の違い、評価方法の違い

[DS\\_202302\\_historical.pdf \(jpma.or.jp\)](https://www.jpma.or.jp/DS_202302_historical.pdf)

## FDAの審査から読み取れること:評価方法について



### 自然歴研究とのCLN2評価尺度の違いに関するFDAからの指摘

#### 1. 評価尺度の違い

190-201/202試験でオリジナルのCLN2評価尺度を改良した尺度を使用  
言語尺度で190-201/202試験の方のスコアが高くなる可能性をFDAが指摘

#### 2. 評価方法の違い

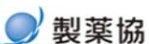
##### 190-201/202試験

撮影されたビデオを4人の医師のうち1人の医師が評価  
同じ患者を常に同じ医師が評価するとは限らなかった

##### 190-901試験

保護者へのインタビューなど後ろ向きな評価が多い  
前向きと後ろ向きな方法で収集されたデータが混在  
同じ被験者でも時点によって収集される方法が異なる場合があった

## スコアの算出方法の違い<sup>[30]</sup>



	Score	190-201/202試験(改良版)	ヒストリカルコントロール(オリジナル)
運動	3	正常に歩行 すなわち際立った運動失調または病的な転倒がない	正常に歩行
	2	独歩は可能(介助なしに10歩以上歩ける)であるが、明らかに不安定な歩き方、頻回に転倒することがある	頻回の転倒、明らかな巧緻運動障害
	1	介助なしには歩行できないまたは四つ這いのみ	介助なしには歩行不能または四つ這いのみ
	0	歩行または四つ這いができない	運動不能、ほぼ寝たきり
言語	3	正常 すなわち明瞭かつ年齢相応の言語を話し、言語能力の低下はまだ認められない	正常
	2	明らかな異常を認める すなわち理解しにくい単語がある ただし、考え、要望、必要なものを短い文で伝えることはできる 本スコアは、これまでの獲得言語(それぞれの患者が最も獲得した到達点)からの低下を意味する	明らかな異常を認める
	1	ほとんど理解不能 すなわち理解できる言葉をほとんど発しない	ほとんど理解不能
	0	理解できる言葉及び発声がない	理解不能または発声なし

[30] BioMarin Pharmaceutical Japan株式会社。(2019).

- オリジナルと改良版の間で一致性を確認する
- 可能な限りHC(ヒストリカルコントロール)と合わせた評価項目を使用する。ただし、前向き試験後にHCを使いたくなった場合は難しい？
- ビデオは改めて第三者による評価が可能である
- 901用の追加症例は評価項目等が適切かもしれない

# セルリポナーゼ アルファ(ブリニューラ®) ディスカッションテーマ②-2

- 外部対照との比較におけるFDAからの指摘事項への対策は？
  - 測定時点の違い
- 評価結果の再現性について考慮すべき対策は？

[DS\\_202302\\_historical.pdf \(jpma.or.jp\)](https://www.jpma.or.jp/DS_202302_historical.pdf)

## FDAの審査から読み取れること:評価方法について



### 3. 評価のタイミングの違い

- 190-201/202試験では評価時点を試験で規定
- 190-901試験では不定期にならざるを得なかった

これら3つの違いから FDAはCLN2評価の再評価を要求

- 190-201/202試験で撮影されたビデオから改めて評価を実施
- 190-201/202試験でもともと得られた結果と比較

## FDAの審査から読み取れること:評価方法について



### 当初の試験結果との一致性<sup>[28]</sup>

比較	運動尺度 重み付きカッパ 係数	言語尺度 重み付きカッパ 係数
190-201/202試験結果 vs 190-201/202 trainerによる評価	0.93	0.82
190-201/202試験結果 vs CLN2 developerによる評価	0.88	0.53
CLN2 developerによる評価 vs 190-201/202 trainerによる評価	0.94	0.56

運動尺度で高い一致性が確認された一方、言語尺度の評価は低い一致性  
FDAは運動尺度のみで試験の評価を実施するよう推奨

・不定期データの補完(関数)も考えられるが、適切でないかもしれない。HCに合わせた試験を計画するのがベストか

# セルリポナーゼ アルファ(ブリニューラ®) ディスカッションテーマ②-3

- 外部対照との比較におけるFDAからの指摘事項への対策は？
  - 患者背景の違い

## FDAの審査から読み取れること:比較可能性について



190-901試験(自然歴研究)と190-201/202試験の患者背景の違い

	190-901試験 (n=42)	190-201/202試験 (n=22)
性別		
男性	25 (60%)	7 (32%)
女性	17 (40%)	15 (68%)
疾患対立遺伝子		
一般的な対立遺伝子の2つがある	24 (57%)	9 (41%)
一般的な対立遺伝子のいずれか一方	11 (26%)	6 (27%)
一般的な対立遺伝子がない	7 (17%)	7 (32%)
出生年代		
1980年より前	4 (10%)	0
1980年代	2 (5%)	0
1990年代	19 (45%)	0
2000年代	16 (38%)	12 (55%)
2010年以降	1 (2%)	10 (45%)

FDAからベースラインの患者背景の違いから有効性評価に影響を与えるのではとの指摘

## FDAの審査から読み取れること:比較可能性について



### マッチング解析

マッチングは1:1で実施  
ベースラインの運動尺度スコア, 年齢(±3か月), 遺伝子変異型  
その結果, 17組の患者が選ばれた。

### 主要評価項目

CLN2評価尺度の**運動尺度のみ**がベースラインから2点以上低下しない, またはベースラインが1点の場合は0点にならない被験者の割合

### 結果

190-201/202試験で, CLN2運動尺度の悪化抑制が認められた  
言語尺度は自然歴研究との比較ができなかったため, 言語尺度の有効性は確立していないと添付文書に記載されている

[DS\\_202302\\_histrical.pdf \(jpma.or.jp\)](https://www.jpma.or.jp/DS_202302_histrical.pdf)

## CLN2運動尺度の結果<sup>[28]</sup>



投与期間	190-201/202試験 (n=17) n (%)	ヒストリカル コントロール (n=17) n (%)	差 % (95%CI)	オッズ比 (95%CI)
48週	16 (94%)	13 (76%)	18% (-19%, 51%)	0.25 (0.005, 2.53)
72週	16 (94%)	11 (65%)	29% (-7%, 61%)	0.17 (0.004, 1.37)
96週	16 (94%)	6 (35%)	59% (24%, 83%)	0.09 (0.002, 0.63)

### 結果

190-201/202試験で, CLN2運動尺度の悪化抑制が認められた  
言語尺度は自然歴研究との比較ができなかったため, **言語尺度の有効性は確立していない**と添付文書(米国)に記載<sup>[31]</sup>

[28] Food and Drug Administration. (2017).

## ディスカッションテーマ②

- ATT推定でよいのか？  
→侵襲性が高いので、実際に治療を受けた人についての評価・推定は妥当である
- 臨床試験およびヒストリカルコントロールに含まれる被験者数が少ない本ケースにおいてなぜマッチングしたのか？臨床試験で収集した22例のうち5例がマッチングから落ちているのは問題である。患者背景で重みを付けて有効性を推定する方法(IPTWなど)なども検討すべきだった
- 臨床研究ではマッチングを用いた解析、層別解析、IPTW、回帰分析などの手法を一通り用いて有効性を推定し、結果が変わらないことを検討することもある
- マッチングから落ちている被験者も多いためマッチングした集団が母集団を適切に反映できていない可能性がある。マッチングした後の患者背景の分布がないので適切にマッチングされたかわからない

セルリポナーゼ アルファ(ブリニューラ®)

## ディスカッションテーマ③-1

### ■ヒストリカルコントロール群設定の要件

DEM-CHILDデータベースに登録されている190-901試験を190-201/202試験の外部対照群として使用する際の要件を満たしているか。また留意点への対応として考えられることはあるか。

一般的にヒストリカルコントロールの利用は重篤な疾患の評価に限定され、以下の(1)から(3)を満たす必要がある。

(1) アンメットメディカルニーズが存在すること

(2) 高い死亡率など、十分に立証された、客観的に測定・検証可能な疾患経過が存在すること

(3) 予想される薬物効果が大きく、自明であり、介入と時間的に密接に関連していること

しかしながら、臨床経過の予測性が高く、客観的に検証可能なアウトカム指標を有する疾患であっても、ヒストリカルデータでは、知られていない、または記録されていない重要な共変量が存在する可能性に留意が必要である。

- ・(3)については作用機序から判断する
- ・E-VALUEによる共変量の評価を実施する

セルリポナーゼ アルファ(ブリニューラ®)

## ディスカッションテーマ③-2

### ■ 外部対照との比較における潜在的バイアスへの対処

同時対照のデータでないため、以下の潜在的バイアスが懸念される。

潜在的バイアスを低減させるために、試験デザインや実施面、解析時に考えられる方策はないか。

#### 選択バイアス

ヒストリカルコントロールと新たな試験データでは、標準的なベースライン特性だけでなく、医療環境、併用療法、標準治療、心理的影響など様々な項目が異なることが想定され、これらが群間の体系的な差異となる。

#### 評価バイアス

盲検化がなされていないため、主観的な評価項目の評価に影響を与える。

- ・ヒストリカルデータに合わせる(ベースライン重症度、試験実施時期)
- ・評価者間の一致性確認、客観性指標も測定する
- ・主観的評価項目の重要性は無視できない(EQ-5D等のアンケート)



# セルリポナーゼ アルファ(ブリニューラ®) ディスカッションテーマ③-3

## ■ 対照群としてのヒストリカルコントロールの利用可能性

DEM-CHILDデータベースに登録されている190-901試験を190-201/202試験の外部対照群として使用することの妥当性について、Pocockの基準(1976)に当てはめてた場合の適切性は？

### ヒストリカルコントロールの利用可能性と潜在的バイアスへの対処



#### ➤ Pocockの基準<sup>[6, 7]</sup>

##### ■ 新規試験の対照群の一部として利用可能な既存のヒストリカルコントロールの条件

1. 詳細に定義された、新規試験の対照治療群と同じ治療を受けている。
2. 新規試験と同じ適格基準を用いて最近実施された臨床試験である。
3. 治療の評価方法が新規試験と同じである。
4. 重要な患者背景の分布が新規試験と同様である。
5. 新規試験とほぼ同じ組織で実施されている。
6. 新規試験との結果の違いを生むと予想される他の要因が存在しない。例えば、新規試験において症例登録スピードが予想よりも上がらないとき、適格基準をみたすぎりぎりの症例が登録されることも起こりうる。そういった場合、新規試験と既存試験で共通の適格基準を用いたとしても登録患者の背景が二つの試験間で異なってくる可能性がある。

➤ Pocock (1976)<sup>[6]</sup>は、すべての条件を満たせばヒストリカルコントロールデータを新規試験の一部として安全に利用できるが、ひとつでも条件が満たされないときは、治療法の比較に大きなバイアスが生じる可能性を否定できないと指摘している

1~6をできるだけ満たすための新規試験デザイン上、及び解析手法での工夫点は？

# セルリポナーゼ アルファ(ブリニューラ<sup>®</sup>) ディスカッションテーマ③

- Pocockの基準は、結局、選択バイアスと交絡を排除したい
- 基準5については、同じ組織でなくても予後に影響する項目を一致させることができればOKである
- RWDの場合、ヒストリカルコントロールとして使用できるだけの質があるか
- ヒストリカルコントロールを用いるうえで、たとえば評価尺度が更新されるなど、時期が異なることによるバイアスの問題が生じる。基準2を満たす最近の試験がある場合は新規にRCTを計画できる可能性もある
- Pocockの基準をすべて満たすような試験及びDBは希少疾患領域では存在しないと思われる。新規試験と患者をマッチングすることが適切かについて、試験計画前にヒストリカルコントロールに用いる試験及びDBを精査すべきである
- たとえ評価項目が同じでも時期が変われば医師の評価方法、評価指針が変わる場合があるので、同じ評価ができるとは限らず、ヒストリカルコントロールとの比較が難しい場合が多いと考えられる
- Pocockの基準をすべて満たす試験、DBは存在しないと思われるため、ヒストリカルコントロールを用いるのであればバイアスは生じる。このバイアスに対して統計家がどう対処するか、解析でどのような工夫が可能か検討する必要がある。バイアスを調整するうえで因果推論的手法が有効だが、どこまで潜在的仮定を受け入れられるかガイダンスがない。ベイズ統計学の利用を含めて発展的手法をどれだけ使えるかPMDAとも検討していく必要がある

# 参考文献

---

- 日本製薬工業協会 (2022). Rare diseaseの治療効果の推測法.  
[https://www.jpma.or.jp/information/evaluation/results/allotment/jtrngf000000085m-att/ds\\_202212\\_rare.pdf](https://www.jpma.or.jp/information/evaluation/results/allotment/jtrngf000000085m-att/ds_202212_rare.pdf)
- 日本製薬工業協会 (2023). 「Rare diseaseの臨床開発における治療効果推測の現状と課題」シンポジウム  
[https://www.jpma.or.jp/information/evaluation/symposium/2023\\_02\\_20.html](https://www.jpma.or.jp/information/evaluation/symposium/2023_02_20.html)
- セロイドリポフスチン症2型治療剤 ブリニューラ®脳室内注射液150mg 投与ガイド  
[https://www.bmrn.co.jp/cms/wp-content/themes/biomarin/pdf/physicians/bmrn\\_brineura\\_04.pdf](https://www.bmrn.co.jp/cms/wp-content/themes/biomarin/pdf/physicians/bmrn_brineura_04.pdf)
- FDA (2019). “Rare Diseases: Common Issues in Drug Development Guidance for Industry”  
<https://www.fda.gov/regulatory-information/search-fda-guidance-documents/rare-diseases-common-issues-drug-development-guidance-industry-0>
- Pocock, Stuart J. "The combination of randomized and historical controls in clinical trials." Journal of chronic diseases 29.3 (1976): 175-188.

事例：ビルトラルセン(ビルテプソ®)

# 概要

対象疾患：デュシェンヌ型筋ジストロフィー（DMD）【[診療ガイドライン \(2014\)](#)】

用法・用量：80mg/kg qw、静脈内投与（日本承認：2020年3月25日）

臨床試験：

– 海外201試験（4～10歳未満の男児16例、日本人患者0例）

• 4週間のプラセボ対照二重盲検期とそれに続く20週間の非盲検期（全例被験薬投与）

• 主要評価項目：ジストロフィン発現量，副次評価項目：運動機能検査〔外部対照（レジストリデータ）と比較〕

– 比較可能性を高めるため、レジストリデータに合わせるように試験デザインを設定

• レジストリデータ（米国）

– 筋ジストロフィーの臨床試験ネットワークが実施したDuchenne natural history study（CINRG DNHS）から65例を選択

• 有効性（運動機能検査）の比較

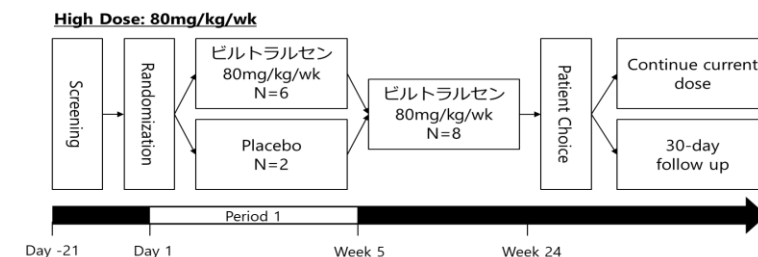
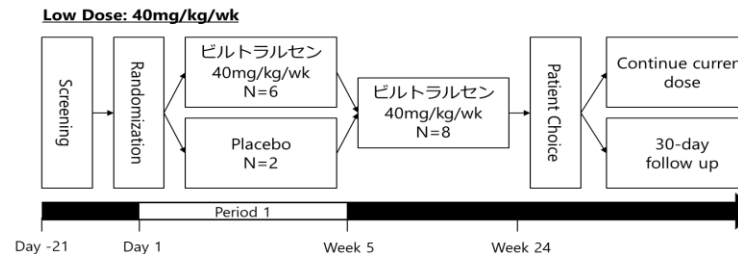
– CINRG DNHS の自然歴群と比較して、運動機能検査の結果で改善が認められた。

• PMDAの外部対照への言及

– ランダム化により比較可能性の担保された集団間での比較ではなく、探索的な検討ではあるものの、得られた試験成績から、本剤投与によるジストロフィン発現増加により、運動機能が改善する傾向は示唆されている。

• 条件付き早期承認制度の利用

– ①安全性/有効性データの早期収集 ②全例対象の使用成績調査 ③安全性/有効性の臨床試験、国内レジストリ調査の実施



[JPMA \(2022\) 図3-1-5-4-1](#)

# 海外201試験の概要（自然歴との比較）

## 201試験（海外試験）

CINRG DNHSと比較可能なデザインを設定

- 運動機能評価の測定時点
- 治験実施施設
- 治験業務の委託先
- 運動機能評価の手順書
- トレーニング方法

試験期間：2016～2018年

目的：
<主要目的>
1. 歩行可能な DMD 男性患者に、低用量（40 mg/kg）及び高用量（80 mg/kg）の NS-065/NCNP-01 を週 1 回静脈内投与した際の安全性及び忍容性を評価する。
2. ウェスタンブロット法による 20-24 週間投与後の筋肉中のジストロフィン発現を指標とし、40 mg/kg 及び 80 mg/kg の NS-065/NCNP-01 を静脈内投与した際の効果を評価する。
3. NS-065/NCNP-01 40 mg/kg 及び 80 mg/kg を静脈内投与した際の薬物動態を評価する。
<副次目的>
1. RT-PCR、免疫蛍光染色及び質量分析（MS）による 20-24 週間投与後の筋肉中のジストロフィン mRNA 及びタンパクの発現を指標とし、40 mg/kg 及び 80 mg/kg の NS-065/NCNP-01 を静脈内投与した際の効果を評価する。
2. NS-065/NCNP-01 40 mg/kg 及び 80 mg/kg を 20-24 週間静脈内投与した際の、立ち上がり時間（TTSTAND）、10m 歩行/走行時間（TTRW）、4 段階段昇り時間（TTCLIMB）、ノース・スター歩行能力評価（NSAA）、6 分間歩行試験（6MWT）及び定量的筋力検査（QMT）による筋力、可動性、運動機能に対する効果を自然歴群（DNHS：Duchenne natural history study）と比較する。
<探索目的>
NS-065/NCNP-01 を 40 mg/kg 及び 80 mg/kg 静脈内投与した際の薬力学を調査する。

<https://www.pmda.go.jp/drugs/2020/P20200408002/index.html>

## CINRG DNHS

CINRG DNHSデータ（男性約440名）より以下の基準で比較対象に適した患者を抽出したところ、男児DMD患者65例が基準を満たし、うち、9例はエクソン53スキッピング対象、56例はそれ以外の患者であった。

データ収集期間：2006～2016年

項目	選択基準
運動機能評価	最低 12 ヶ月の時間機能検査を実施済み
年齢	4 歳以上 10 歳未満
地域	米国又はカナダ
ステロイド投与	最低 3 ヶ月間継続的に服用
治験参加	他のエクソンスキッピング薬の臨床試験には非参加
遺伝子検査	検査結果が判明している
除外対象	プロモーターからエクソン 8 の間に変異のある患者、エクソン 44 スキッピング治療対象患者 インフレーション変異を持つ患者、及びリーディングフレームかどうか判断できない変異を持つ患者

[https://www.info.pmda.go.jp/go/interview/1/530263\\_1900400A1025\\_1\\_003\\_1F.pdf](https://www.info.pmda.go.jp/go/interview/1/530263_1900400A1025_1_003_1F.pdf)

# 海外201試験の試験結果

## 主要評価項目の結果(ビルトラルセン)



ウェスタンブロット法によるジストロフィン発現のベースラインからの変化<sup>[34]</sup>

投与群	40 mg/kg 群	80 mg/kg 群
評価例数	8	8
ミオシン重鎖による標準化(%)		
ベースライン	0.3 ± 0.10	0.6 ± 0.82
ベースラインからの変化	5.4 ± 2.40	5.3 ± 4.48
P値*	0.0004	0.0123
α-アクチニンによる標準化(%)		
ベースライン	0.2 ± 0.22	0.4 ± 0.67
ベースラインからの変化	5.2 ± 2.83	3.3 ± 2.47
P値*	0.0012	0.0074

\* P値は対応のあるt検定を用いて算出

40 mg/kg群、80 mg/kg群で改善の傾向がみられた

[34] 医薬品医療機器総合機構. (2020).

55

## 運動機能評価の結果(ビルトラルセン)



ヒストリカルコントロールを用いた立ち上がり時間の解析結果<sup>[37]</sup>

	201試験			ヒストリカルコントロール
	40 mg/kg	80 mg/kg	全体	全体
評価例数	8	8	16	65
ベースライン	4.17 ± 1.146	4.76 ± 2.580	4.44 ± 1.956	5.55 ± 3.041
変化量(25週)	0.05 ± 1.446	-0.44 ± 0.750	-0.19 ± 1.141	0.66 ± 1.845
P値*	0.2077	0.0543	0.0366	

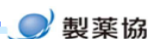
\* P値はmixed model for repeated measuresを用いて算出。

立ち上がり時間なので変化量がマイナスであることが改善を意味  
ビルトラルセンが良好である結果が得られた

[37] 日本新薬株式会社. (2019)

56

## PMDAからのコメント(ビルトラルセン)



審査報告書<sup>[34]</sup>より

海外201試験と CINRGの自然歴集団等の外部対照との比較について、ランダム化により比較可能性の担保された集団間での比較ではなく、探索的な検討ではあるものの、得られた試験成績から、本剤投与によるジストロフィン発現増加により、運動機能が改善する傾向は示唆されている

PMDAから肯定的な評価を受けている

[34] 医薬品医療機器総合機構. (2020).

57

## 規制当局の反応が異なる要因として考えられるもの



事前にヒストリカルコントロールを利用するよう試験がデザインされていたか

以下、審査報告書<sup>[34]</sup>の記載を抜粋

海外201試験の運動機能評価の測定時点、試験実施施設、試験業務の委託先、運動機能評価の手順書、及びトレーニング方法は、CINRG自然歴集団と比較可能なデザインとなるように設定した。

比較可能性を確保する努力が、規制当局に前向きに受け入れてもらえた要因であると推察できる

[34] 医薬品医療機器総合機構. (2020).

58

# 希少疾病医薬品開発に関連するガイダンス

## Rare Diseases: Common Issues in Drug Development (FDA, 2019) [7]



### Rare Diseases: Common Issues in Drug Development Guidance for Industry

#### DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 60 days of publication in the Federal Register of the notice announcing the availability of the draft guidance. Submit electronic comments to <https://www.regulations.gov>. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, P.O. Box 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the Federal Register.

For questions regarding this draft document, contact (CDER) Lucas Kempf at 301-796-1140 or (CBER) Office of Communication, Outreach, and Development at 800-835-4709 or 240-402-8010.

U.S. Department of Health and Human Services  
Food and Drug Administration  
Center for Drug Evaluation and Research (CDER)  
Center for Biologics Evaluation and Research (CBER)

January 2019  
Rare Diseases  
Revision 1

**【ガイダンスの目的】** 希少疾病の治療/予防を目的とする医薬品/生物学的製剤を開発する治験依頼者が、より効率的な医薬品開発プログラムを推進するための支援

#### 重要とされているポイント

- 疾患の自然歴に関する十分な理解
- 疾患の病態生理, 及び薬剤の作用機序に関する十分な理解
- 臨床試験による治療の安全性を裏付けるための毒性に関する考察
- 評価項目の選択または開発
- 安全性及び有効性を確立するためのエビデンス
- 医薬品開発中における製造上の留意事項
- 患者, 介護者, 及び支援者の開発プログラムへの参加
- 規制当局との対話

8

## Duchenne Muscular Dystrophy and Related Dystrophinopathies: Developing Drugs for Treatment Guidance for Industry

U.S. Department of Health and Human Services  
Food and Drug Administration  
Center for Drug Evaluation and Research (CDER)  
Center for Biologics Evaluation and Research (CBER)

February 2018  
Clinical/Medical

<https://www.fda.gov/media/92233/download>

## Rare Diseases: Common Issues in Drug Development (FDA, 2019) [7]



### 【安全性及び有効性を確立するためのエビデンス】 ヒストリカルコントロール群の利用

- 医療上の必要性が満たされていない重篤な希少疾病に対しては、同時対照群へのランダム化を伴わない、ヒストリカルコントロール群の利用に関心が示されることが多い。しかし、**同時対照群でないために生じる系統的な差を排除できないことが、ヒストリカルコントロール群を用いるデザインの大きな問題である。**一般的にヒストリカルコントロールの利用は重篤な疾患の評価に限定され、以下の(1)から(3)を満たす必要がある。

- (1) アンメットメディカルニーズが存在すること
- (2) 高い死亡率など、十分に立証された、客観的に測定・検証可能な疾患経過が存在すること
- (3) 予想される薬物効果が大きく、自明であり、介入と時間的に密接に関連していること

- しかしながら、臨床経過の予測性が高く、客観的に検証可能なアウトカム指標を有する疾患であっても、ヒストリカルデータでは、知られていない、または記録されていない重要な共変量が存在する可能性に留意が必要である。

9



# ディスカッションテーマ①

---

## 海外201試験をランダム化比較試験で実施しなかった理由は何か？

- 対象疾患：デュシェンヌ型筋ジストロフィー（DMD）
- 病因及び患者層：
  - デュシェンヌ型筋ジストロフィーは、ジストロフィンと呼ばれる細胞膜裏打ち蛋白質の欠損によって発症しており、男児3,500人に1人が発症するといわれている。運動発達の遅れによって2～3歳に病気と診断され、12歳までには車椅子での生活となり、14歳を過ぎると1/3の患者に心筋障害を生じる<sup>1</sup>。
- 対象集団：エクソン53スキッピングにより治療可能なジストロフィン遺伝子の欠失が確認されている男性DMD患者
  - 選択基準（一部）
    - ベースライン時に4歳以上10歳未満、北アメリカ（米国又はカナダ）
    - ベースライン時にコルチコステロイドを少なくとも3カ月間投与しており、12～24カ月の観察期間を通して継続して使用している。
    - 遺伝子検査の結果、重複変異、ナンセンス変異又はフレームシフトを引き起こす変異をもつ患者も含める。
- 評価項目：
  - 主要評価項目：筋生検におけるジストロフィンタンパク質の発現（ウェスタンブロット法）
  - 副次評価項目（運動機能評価）：
    - 時間機能検査（10 m歩行/走行時間、床からの立ち上がり時間、6分間歩行距離、4段階昇り時間、ノース・スター歩行能力評価）
    - 定量的筋力検査

<sup>1</sup> 戸田達史. "2. 筋ジストロフィーの分子機構と治療戦略." 日本内科学会雑誌 105.9 (2016): 1578-1587.

## ディスカッションテーマ①

---

### 海外201試験をランダム化比較試験で実施しなかった理由は何か？

- 倫理的な懸念からプラセボを対照として置くことは避けたい。
  - 本剤の対象疾患に対する他の治療薬がない。
  - 海外201試験では、対症療法（全身性コルチコステロイド剤）の併用を可能としている。
  - eteplirsenが承認されている。
- 探索的試験の場合、本剤を投与した時の情報（有効性など）を収集することが目的である。
  - 検証的試験の場合は検討？
- 被験者登録が困難である。疾患の自然経過がわかっている。患者の多くが小児であるため、治療を早期に開始した方が良い。
- 評価項目が自覚症状ではなく、プラセボを置かなくても評価しやすい。
- 客観的な主要評価項目であるので、2例4週間のプラセボ群の推移をもってある程度自然経過を判断できる。
- 海外201試験でプラセボを置いたのは安全性の評価のためでは？
- 対象集団が小規模の場合、ランダム化の利点が維持できているのか？
  - 交絡が制御されているのか？
  - 因果的な影響は制御されているが、データのばらつきにより介入効果を精度の良く推測が困難になる？
  - 動的割り付けを行うなど、効率よく群間のバランスをとる方法の利用が考えられる。
  - 介入情報を収集することを主要な目的にする場合、対照を置くより、介入情報を多く収集することに主眼を置くのでは？

## ディスカッションテーマ②

---

CINRG DNHSのレジストリデータを海外201試験の外部対照群として使用することの妥当性について Pocockの基準（1976）に当てはめてた場合適切だったのか？またこの基準に外部対照群を評価するときの課題は？

1. 詳細に定義された、新規試験の対照治療群と同じ治療を受けている.
2. 新規試験と同じ適格基準を用いた最近実施された臨床試験である.
3. 治療の評価方法が新規試験と同じである.
4. 重要な患者背景の分布が新規試験と同様である.
5. 新規試験とほぼ同じ組織で実施されている.
6. 新規試験との結果の違いを生むと予想される他の要因が存在しない. 例えば、新規試験において症例登録スピードが予想よりも上がらないとき、適格基準をみたすぎりぎりの症例が登録されることも起こりうる. そういった場合、新規試験と既存試験で共通の適格基準を用いたとしても登録患者の背景が二つの試験間で異なってくる可能性がある.

[武田ら\(2015\)計量生物学](#)

## ディスカッションテーマ② (1/4)

CINRG DNHSのレジストリデータを海外201試験の外部対照群として使用することの妥当性についてPocockの基準（1976）に当てはめてた場合適切だったのか？またこの基準に外部対照群を評価するときの課題は？

### 全般

- 患者背景の類似性評価が課題になるだろう（海外201試験はCINRG DNHSに合わせることを行った）。
- 対象疾患に対する予後因子の特定が重要になる。
- レジストリデータの場合、評価の方法や項目を揃えにくい（外部データは限られた情報を収集、異なる評価間隔など）。
- 一部の基準は、データを収集してからしか評価ができない場合もある。
- データが得られたあとで、基準に合致しているかを確認しなければいけない場合もあり、事前に保証できるものではない。
- すべての基準を満たすのは難しく、満たさなかった場合は何らかの工夫が必要になる。被験者数が少数だと限界がある。
- 統計学的な手法（共変量の調整、傾向スコアなど）を用いて、バランスをとることも必要になる。
- 多角的に評価をしたときの乖離の影響を考察できるか。
- どの時点で外部対照データが開示されるのか（臨床試験開始前？開始中？）、同じ外部対照データが他の臨床試験で使用されているのかも重要になるのでは？
- ビルトラルセンの場合、基準3~5は満たしているのでは？
- 1976年に提案されたPocockの基準はレジストリデータを想定していないのでは？

## ディスカッションテーマ② (2/4)

---

CINRG DNHSのレジストリデータを海外201試験の外部対照群として使用することの妥当性についてPocockの基準（1976）に当てはめてた場合適切だったのか？またこの基準に外部対照群を評価するときの課題は？

基準 1：詳細に定義された、新規試験の対照治療群と同じ治療を受けている

- 本事例が詳細に定義されていたかは不明だが、プラセボの条件は満たしている。ステロイド投与はどうか？併用薬も揃えられると望ましい→適格基準に少なくとも3か月のグルココルチコイドの用量を利用しているという基準が入っている。

基準 2：新規試験と同じ適格基準を用いた最近実施された臨床試験である

- データソースによっては、適格基準を揃えることが課題になるのでは？
  - レジストリだと適格基準は広めにとるので、臨床試験と同じは困難。
  - 本事例は、絞る規準を事前に決め、バイアスを最小にする努力をした。レジストリから対象集団を特定できるかが重要だが、それができていた。
- 本事例の場合、レジストリデータ収集期間と臨床試験実施時期が重なっていない。←最近実施には当てはまらない？

## ディスカッションテーマ② (3/4)

---

CINRG DNHSのレジストリデータを海外201試験の外部対照群として使用することの妥当性について Pocockの基準 (1976) に当てはめてた場合適切だったのか？ またこの基準に外部対照群を評価するときの課題は？

基準 3 : 治療の評価方法は新規試験と同様である。

- 収集期間が重なっていなくても、ステロイド使用などが同じであれば同等とみなしてよいのでは？
- 本事例の場合、評価方法は同じであるか？ → 運動機能の評価時点, 実施施設, 試験業務の委託先, 運動機能評価の手順書, 及びトレーニング方法は, CINRGの自然歴集団と比較可能なデザインとなるように設定した。
- 外部対照との比較が客観的指標ではない運動機能の改善。主要と副次の相関評価などの橋渡しが必要では？ → 代替エンドポイントの利用 (希少疾患では難しい?) → バイオマーカーと運動機能との関連は臨床試験でなくても評価できる? 先行研究がある? → ジストロフィン発現の検査は侵襲性が高いので関連性の評価は不明。ジストロフィン発現を下げたら運動機能が改善するかはたぶんみられていない。 → 主要評価項目がジストロフィン発現になったのは、生物学的妥当性と、プラセボ効果を受けにくい客観的指標であるため。

## ディスカッションテーマ② (4/4)

---

CINRG DNHSのレジストリデータを海外201試験の外部対照群として使用することの妥当性についてPocockの基準（1976）に当てはめてた場合適切だったのか？またこの基準に外部対照群を評価するときの課題は？

基準 4：重要な患者背景の分布が新規試験と同様である

- 類似性をどのように判断するか。少数の被験者数では限界がある。
- ヒストリカルデータの患者背景は偏っているのは、臨床試験とどうそろえるのか。→立ち上がり時間、10m歩行はあまり変わらない

基準 5：新規試験とほぼ同じ組織で実施されている

- 本事例では、この基準を満たしていないのでは？
- 最低限基準を満たす必要がある項目はどれか含めて課題（疾患による）。
- 分布が類似とはどのように判断するか？

## ディスカッションテーマ③

海外201試験の有効性評価は、主要評価項目の評価においては対照群との比較を行わず、副次評価項目において外部対照群との比較を行っている。このような解析計画から得られた結果の考察における留意点は？

- 主要評価項目（ジストロフィン蛋白質の発現）及び副次評価項目（運動機能検査）について、各用量群の被験者内変化を対応のあるt検定、2用量群間の比較では対応のないt検定
- 運動機能検査の結果は、CINRG DNHSの患者集団（65名）の結果とモデル解析（MMRM）より比較
- 検定はいずれも有意水準5%（両側）に基づき実施。多重性の調整は行っていない。

主要評価項目の結果(ビルトラルセン)

ウエスタンブロット法によるジストロフィン発現のベースラインからの変化<sup>[34]</sup>

投与群	40 mg/kg 群	80 mg/kg 群
評価例数	8	8
ミオシン重鎖による標準化(%)		
ベースライン	0.3 ± 0.10	0.6 ± 0.82
ベースラインからの変化	5.4 ± 2.40	5.3 ± 4.48
P値*	0.0004	0.0123
α-アクチニンによる標準化(%)		
ベースライン	0.2 ± 0.22	0.4 ± 0.67
ベースラインからの変化	5.2 ± 2.83	3.3 ± 2.47
P値*	0.0012	0.0074

\* P値は対応のあるt検定を用いて算出

40 mg/kg群、80 mg/kg群で改善の傾向がみられた

[34] 医薬品医療機器総合機構. (2020).

55

運動機能評価の結果(ビルトラルセン)

ヒストリカルコントロールを用いた立ち上がり時間の解析結果<sup>[37]</sup>

評価例数	201試験			ヒストリカルコントロール
	40 mg/kg	80 mg/kg	全体	全体
ベースライン	4.17 ± 1.146	4.76 ± 2.580	4.44 ± 1.956	5.55 ± 3.041
変化量(25週)	0.05 ± 1.446	-0.44 ± 0.750	-0.19 ± 1.141	0.66 ± 1.845
P値*	0.2077	0.0543	0.0366	

\* P値はmixed model for repeated measuresを用いて算出。

立ち上がり時間なので変化量がマイナスであることが改善を意味  
ビルトラルセンが良好である結果が得られた

[37] 日本新薬株式会社. (2019)

56

[https://www.jpma.or.jp/information/evaluation/symposium/jtrngf0000001cap-att/DS\\_202302\\_histrical.pdf](https://www.jpma.or.jp/information/evaluation/symposium/jtrngf0000001cap-att/DS_202302_histrical.pdf)



## ディスカッションテーマ③

海外201試験の有効性評価は、主要評価項目の評価においては対照群との比較を行わず、副次評価項目において外部対照群との比較を行っている。このような解析計画から得られた結果の考察における留意点は？

- 副次評価項目（運動機能）の評価のみ外部対照と比較したことは明確にする必要がある。
  - 客観的指標ではなく、プラセボ効果を考慮する必要がある。
- 検証的試験ではないため、多重性を考慮する必要もなく、主要な評価は投与の前後差行い、疾患メカニズムで考察をすることになる。
- 感度分析を実施する必要がある。
- ランダム化できていない→交絡が存在する可能性がある。比較可能性はどの程度保てるのか。
- 主要評価項目がベースラインからの比較。プラセボ効果が存在する場合、有効な評価ができていないか疑問。
- MMRMの仮定は妥当か。→被験者数が少ないので、モデルの評価が難しい。
- 主要評価項目（ベースラインからの変化量）と副次評価項目（群間比較）の意味合いが違う。
- 仮説に基づいて例数設計したわけではない。P値にどれほどの意味があるか。変化量の差の臨床的な意味づけなど、P値以外の議論が必要。→点推定値が逆転している低用量と高用量を統合してP値が0.05以下であることにどれほど意味があるのか。→5%を超えるか超えないかがこの疾患の閾値
- 主要と副次で矛盾する結果が出た場合、外部対照が入ることにより解釈が困難になる。

# ディスカッションテーマ④ - FDAガイダンス（2018） -

---

## 試験デザイン

FDAは、有効性を示す最も効率的な方法としてランダム化プラセボ対照試験を強く推奨している。しかし、状況によっては、外部対照を用いた試験（ヒストリカルコントロール試験）を、承認を裏付ける有効性のエビデンスとする可能性がある適切かつ十分にコントロールされた試験とみなす場合がある。治験依頼者は、試験の効率を高めるデザイン（アダプティブデザインなど）を使用することができる。

## 試験対象集団

複数の疾患ステージ及び表現型にわたって安全性及び有効性を特徴付ける必要があるが、治験依頼者は科学的に正当化された場合、特定された疾患サブグループ（例えば、特定のジストロフィン変異を対象とする薬剤）を対象とした開発をすることができる。

臨床症状の悪化を遅らせる可能性があるが、既存の筋機能障害の改善又は回復が期待できない薬剤については、予後的エンリッチメントを考慮することが有用となりうる。

## 有効性評価項目

FDAは治験依頼者に対し、幅広い症状及び疾患ステージを有する患者を正当かつ正確に評価できる評価項目を提案し、必要に応じて開発するよう奨励する。治験依頼者は、有効性評価項目の選定又は開発の早い段階でFDAと協議する必要がある。

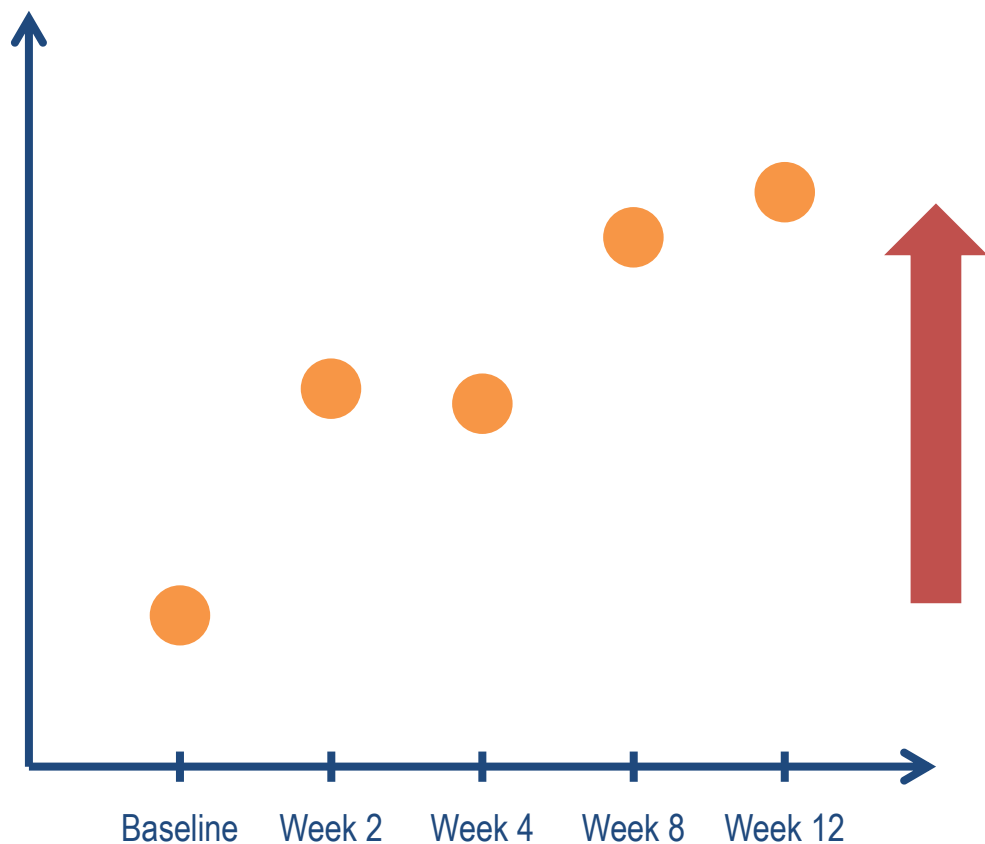
## 統計的考察

連続量の評価項目については、治験依頼者は一般的に各投与群のベースラインからの変化について統計解析を行い、1つ以上の特定の時点で被験薬群と対照薬群の平均変化を比較して治療効果を評価する。治験依頼者は、傾き（又は変化率）によるベースラインからの変化を評価する方法も検討できる。

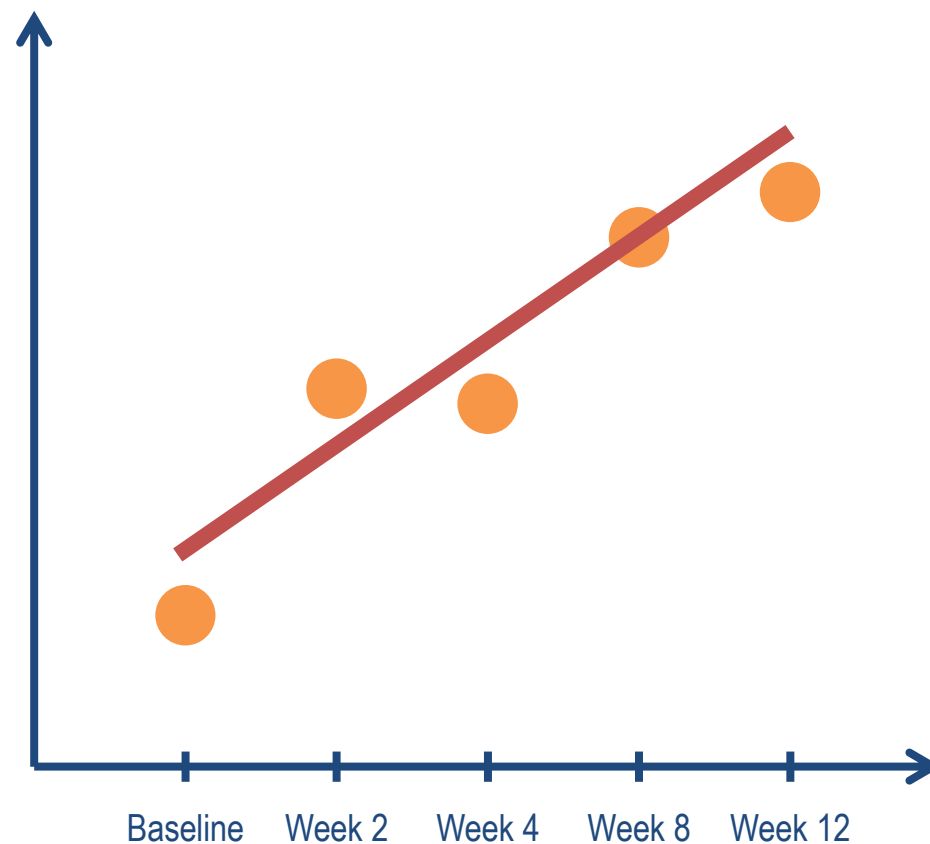
## ディスカッションテーマ④ - 統計的考察 -

---

ベースラインから12週時までの変化量



ベースラインから12週時までの変化の傾き



## ディスカッションテーマ④

---

FDAガイダンスの内容を踏まえ、以下の観点で自由に議論しましょう。

### ① 試験デザイン

–ランダム化比較試験（RCT）、外部対照試験、アダプティブデザインなど

### ② 対象集団の選択

–広範囲な患者背景（年齢や病態など）を持つ患者集団、エンリッチメント戦略（予後エンリッチメント、予測エンリッチメント）

### ③ 評価項目の選択

–評価項目の開発、代替評価項目の設定など

### ④ 統計解析

–「変化量 vs. 傾き」：解析結果の解釈、サンプルサイズなど

## ディスカッションテーマ④ (1/2)

FDAガイダンスの内容を踏まえ、以下の観点で自由に議論しましょう。

### ① 試験デザイン

- 臨床的観点を踏まえてデザインを検討する。
- プラセボ対照試験（規模に関係なく、検証的試験では必須）。
  - プラセボを置くことが難しい場合は、外部対照試験（単群試験又はプラセボ群の一部）を実施する。
  - 外部対照試験の場合は、利用可能なヒストリカルデータを事前に検討する必要がある。
- 海外201試験はプラセボを4週間投与。その程度であれば許容されるので、評価項目次第である。4週間投与で判断できる評価項目を見つけられるのであれば、プラセボ対照も許容できる。→ 4週間投与でできたとしても、被験者数は課題。今回の試験でも被験者数が限られている。
- プラセボをどの程度投与できるかの課題もある。基本的には外部対照。
- プラセボ + 外部対照が好ましい。今回の試験で可能か？
- 難病の場合はレジストリを使用して外部対照が良いのでは？ → レジストリの中でランダムサンプリングをして臨床試験を行い、参加しなかった人で適格基準を満たす人を対照とする。

### ② 対象集団の選択

- エンリッチメント試験は薬効が期待できる対象集団に絞る試験デザイン。希少疾患にはあてはまりづらい。予後的エンリッチメント（イベントを起こしそうな人を登録する）は今回の場合は歩行時間が一定以上の方に限定することになる。治療効果に関しては、ターゲット遺伝子も決まっていることからプラスアルファは難しい。

## ディスカッションテーマ④ (2/2)

FDAガイダンスの内容を踏まえ、以下の観点で自由に議論しましょう。

### ③ 評価項目の選択

- 主要評価項目と副次評価項目をどちらにするかは、臨床試験でどの程度のエビデンスを確保できるかに依存する。プラセボ効果を排除できたり、客観的に評価できる指標を主要評価項目としたのではないか。→主要評価項目は診断ガイドラインに入っているのが妥当ではないか。外部対照をメインで考えるか有効性が決められているもので評価するかが重要。→ジストロフィンと運動機能との関連は試験開始段階では明確にされていないので、診断基準のマーカーなので妥当であるとは一概には言えない。そのため副次評価項目との合わせ技という方針にした。

### ④ 統計解析

- バイズ流の統計手法の利用を検討すべきか？、又はバイズ流の統計手法ありきで計画した方が良いか？ 採用するにはそれなりのマンパワー（シミュレーションや結果の解釈など）が必要になるのでは？
- ベースラインでの調整は必要ないか。→ベースラインはそれほど変わらない。対応のあるt検定よりも共分散分析の方が効率が上がるのではないか。
- 変化が非線形の場合に傾きは適切ではないのでは。24週時点での変化は重要か。
- 自然経過は分かっている、治療反応が分かっているわけではないので、1時点での変化量が解釈しやすいのではないか。
- イベント発現までの生存時間解析もできたのではないか→試験期間が延びてしまう。歩行不能は10歳くらいなので5年とかかかってしまう。被験者数も少ない。妥当な評価はできそうだが。検出力の観点で不安が残る。連続変数の方が検出力は確保できる。
- 自然歴の副次は組み入れ基準を絞っているとのことだが、マッチングの方が感度が良くなるのではないか。→過度にマッチングすると被験者がいなくなるのでキャリパーを設けては。マッチしない場合もある。（test-then-pool）

## その他

---

- 探索的枠組み→製造販売後の情報が重要
  - 特定のイベント発生までの時間をヒストリカルデータと比較→起点の設定が難しい
  - 有害事象が10%以上発現のものしか臨床試験ではとれないので、製造販売後も安全性管理情報の収集が必要
    - 希少疾患においてレジストリは重要である。レジストリの枠の中で評価できるのでは？全例調査よりも効率的？

## 参考文献

---

- [日本製薬工業協会 \(2022\). Rare diseaseの治療効果の推測法](#)
- [日本製薬工業協会 \(2023\). 「Rare diseaseの臨床開発における治療効果推測の現状と課題」シンポジウム](#)
- [日本神経学会、日本小児神経学会、国立精神・神経医療研究センター 監修 \(2014\). 「デュシェンヌ型筋ジストロフィー診療ガイドライン2014」](#)
- [日本新薬株式会社 \(2020\). ビルテプソ点滴静注250 mgに関する資料](#)
- [日本新薬株式会社 \(2020\). ビルテプソ インタビューフォーム](#)
- Pocock, Stuart J. "The combination of randomized and historical controls in clinical trials." *Journal of chronic diseases* 29.3 (1976): 175-188.
- [Food and Drug Administration \(2018\) "Duchenne Muscular Dystrophy and Related Dystrophinopathies: Developing Drugs for Treatment - Guidance for Industry."](#)
- 武田健太郎ら (2015) . "臨床試験におけるヒストリカルコントロールデータの利用." *計量生物学* 36.1 25-50



# 事例 : Eteplirsen

Eteplirsenに関する情報は、以下の資料から抜粋しております。

- 日本製薬工業協会 (2022). Rare diseaseの治療効果の推測法.
- 日本製薬工業協会 (2023). 「Rare diseaseの臨床開発における治療効果推測の現状と課題」シンポジウム

# 対象疾患

---

## DMD(デュシェンヌ型筋ジストロフィー)

- 骨格筋繊維の変性・壊死と不完全再生のサイクルを繰り返しながら間質の繊維化・脂肪化が進行する遺伝性疾患群
- 進行性の骨格筋萎縮と筋力低下、日常生活動作が低下
- 呼吸筋不全や心不全といった重篤な合併症を併発することもある難病
- ジストロフィンと呼ばれる細胞膜裏打ち蛋白質の欠損によって発症
- 12歳までには車椅子での生活
- 14歳を過ぎると1/3の患者に心筋障害を生じる

# Eteplirsen

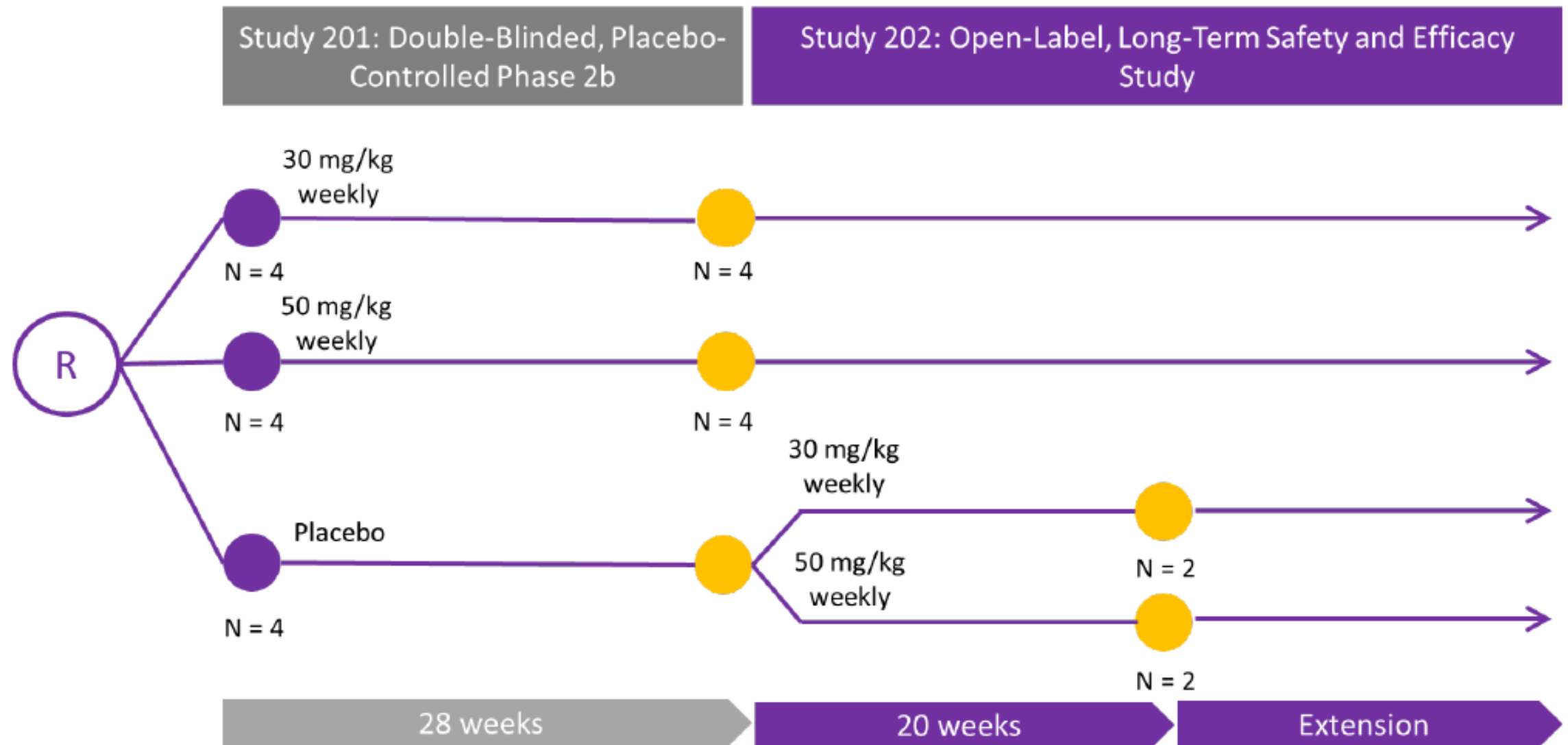
---

- エクソン・スキップ治療と呼ばれる治療メカニズムに基づく薬剤
- ジストロフィン蛋白発現を回復させることによって、デュシェンヌ型筋ジストロフィーの症状悪化を緩和させる薬剤

## 承認状況

- 2016年9月米国で承認
- 日本では未承認

# 2つの臨床試験



# 2つの臨床試験

---

## 201試験

- ランダム化二重盲検プラセボ対照試験, 米国1施設
- プラセボ群, 30mg/kg 群, 50mg/kg 群で各4名
- 主要評価項目: 20週時点の筋肉生検組織中のジストロフィン陽性繊維の割合(%), **薬理的な代替評価項目**

## 202試験

- オープンラベル, 201試験からの継続長期投与試験
- 主要評価項目: 6分間歩行の距離(m)
- 201試験終了後, 12名の被験者全員が移行
- 48週時点の評価に**疾患レジストリデータを外部対照として利用**

# 201試験の結果

時点		プラセボ群 N = 4	30 mg/kg 群 N = 4	50 mg/kg 群 N = 4
Baseline	Mean	15.64	18.19	11.00
	Median	15.58	17.80	11.51
	SD (SE)	10.742 (5.371)	5.501 (2.751)	4.668 (2.334)
	Min, Max	3.2, 28.2	11.9, 25.3	5.4, 15.6
On-Treatment	Mean	11.59	41.14	11.79
	Median	9.44	38.77	11.81
	SD (SE)	7.130 (3.565)	10.097 (5.049)	4.456 (2.228)
	Min, Max	5.7, 21.7	32.7, 54.3	6.4, 17.2
Change from Baseline	Mean	-4.05	22.95	0.79
	Median	-6.13	23.46	2.52
	SD (SE)	5.834 (2.917)	5.792 (2.896)	7.099 (3.549)
	Min, Max	-8.5, 4.5	15.9, 29.0	-9.3, 7.4
	p-value*		0.002	0.958

\* 共分散分析モデルに基づく。モデルにおいては、順序データとしての投与群（プラセボ群，30mg/kg 群，50mg/kg 群）と固定効果とし、ベースライン値とデュシェンヌ型筋ジストロフィーと診断されてからの期間を共変量とした。

## 主要評価項目

- 20週時点の筋肉生検組織中のジストロフィン陽性繊維の割合（%）

## 結果

- 30 mg/kg群は22.95%の増加、50 mg/kg群は、0.79%の増加

# 202試験：外部対照の選択

---

- Eteplirsen投与例の観察後に2つの患者レジストリから作成。  
Eteplirsen投与例とマッチさせるために、以下の条件で患者の選択。
  1. ベースライン時点でのコルチコステロイドの使用
  2. 6分間歩行試験に関する十分な経時データの存在
  3. 7歳以上の年齢
  4. あらゆるエクソン・スキップ治療が適用できる遺伝子型
  5. エクソン51に関するスキップ治療が適用できる遺伝子型
- 患者選択の結果、186名のレジストリデータから、13名の患者をヒストリカルコントロール群として選択

# 202試験の結果

群		年齢	6 分間歩行試験 (ベースライン)	6 分間歩行試験 (36 か月後)**
ヒストリカルコントロール	N	13	13	13
	Mean / LS Mean* (SE)	9.45 (0.403)	357.6 (18.51)	115.1 (33.54)
	Min, Max	7.3, 11.8	200, 458	
eteplirsen	N	12	12	12
	Mean / LS Mean* (SE)	9.41 (0.342)	363.2 (12.18)	256.4 (33.11)
	Min, Max	7.3, 11.0	256, 416	

\* LS Mean は 6 分間歩行試験 (36 か月後) に対して算出されている。

\*\* LS Mean の群間差 (36 か月後) は 141m であり, p 値は 0.009 である。

## 主要評価項目

- 6分間歩行の距離 (m)

## 結果

- LS Meanの群間差は141m, 群間差の検定のp値は0.009



# FDA審査コメント①

---

- ICH E10ガイドライン「臨床試験における対照群の選択とそれに関連する諸問題」では「外部対照試験を採用すべきか考慮するのは、一般に、被験治療が全ての既存の治療法より優れているとの事前の確信がきわめて強い...」と記載があるが、この条件はeteplirsenには該当しない。
- ICH E10ガイドラインでは「外部対照を採用するのは、エンドポイントが客観的であり」との記載があるが、6分間歩行試験はモチベーションによる影響を受ける。

# FDA審査コメント②

---

- ヒストリカルコントロールの受け入れについて, Pocockは「治療の評価方法が新規試験と同じである」, 及び「新規試験とほぼ同じ組織で実施されている」という基準を挙げている. これは, 6分間歩行試験のようなエンドポイントの評価においては特に重要であるが, この要件は満たされていない.
- 更に重要な点として, 今回の申請において, ヒストリカルコントロール群は事後的に特定された点が挙げられる. そのため, 定量化できない選択バイアスがもたらされた可能性がある. もしヒストリカルコントロールを用いるのであれば, 対照群の選定や, 選択基準の設定は, 実薬群と対照群のアウトカムを知らない状態で事前に計画されるべきであった. また, ヒストリカルコントロールを用いた202試験の成功基準は, 予め治験実施計画書で規定されていなかった.

# テーブルディスカッションの論点

## 1. 外部対照の根拠

ICH E10は外部対照について「被験治療が全ての既存の治療法より優れているとの事前の確信がきわめて強い」場合と述べている。この事例では、事前に得られている情報がほとんどなかった。どのような情報を、どのように解釈すると、効果を事前に類推できるか？

- どのような事前情報があり得そうか？
  - ジストロフィン陽性繊維改善によって、症状改善が見られるという仮説が事前の研究論文(既存論文, 非臨床など)で証明されていれば, 説明できるかもしれない。
    - 新規の作用機序であれば, 非臨床データに頼ることになるのでは...
  - Primary Endpointにかかわる情報として, 参考となる非臨床のデータなどが必要になるのではないか。
    - 臨床試験を計画する段階で, 必要な非臨床データが得られているかは課題？
    - Primary Endpointの効果の情報も必要だが, 評価項目として妥当か? という情報も必要かもしれない
  - できる範囲では, 非臨床データ. もしくはPh1(健康成人?)で有効性の評価項目についても確認しておき, 動くかどうか確認することも一案ではないか。
    - 健康成人試験が成立するのであれば, 評価し作用のあたりをつけておくのはアリではないか
  - 前臨床, ラボレベルの評価で効果が見られているのであれば, 臨床でもそれ以上の効果が見られるという仮定を置いてシミュレーションを実施し, それを事前情報とする案
- 臨床症状を評価項目にしているので, 臨床医に相談の上「閾値」を事前に検討しておくことも一案ではないか。それが効果を類推する指標の1つになるのでは? (プラセボ群との群間差ではなく, 単群での評価を考えることも一案では)
  - プラセボ群で動かない, という自然歴の情報があればこの考えは実装できたのではないか。
- 医療機器では, 介入した結果効果が見られることが自明であるため, 単群での評価を採用することもある

# テーブルディスカッションの論点

## 1. 外部対照の根拠

ICH E10は外部対照について「被験治療が全ての既存の治療法より優れているとの事前の確信がきわめて強い」場合と述べている。この事例では、事前に得られている情報がほとんどなかった。どのような情報を、どのように解釈すると、効果を事前に類推できるか？

- 単群、閾値を置いた評価を実施するのであれば、疾患の理解と自然歴としてどの程度動くのか評価する必要がある。どのように評価するのが良いか？
  - 自然歴のデータをとり、その後評価フェーズに移る計画を実施したことがある。
- 効果の観点
  - 作用機序にかかわる情報を、非臨床からもってきて、ジストロフィンが改善すれば症状が改善することを示す
  - 臨床医と相談の旨、閾値を設定
- Primary Endpoint設定の観点
  - 評価項目として設定した内容が妥当である、という外部情報
- 自然歴研究やレジストリにより、Natural Historyを知ることが重要
- 既存の治療法に関する研究成果を調査する
- バイオマーカーと予後の関係性、薬理的な根拠
- 201試験の結果におうじて、202試験の計画に柔軟性をもたせる
- “被験治療が全ての既存の治療法より優れているとの事前の確信がきわめて強い”という条件は被検者保護の観点なのか解釈上の必要条件なのか？
- 薬理的な考察が重要になる可能性があるので、統計担当者だけの根拠づけは難しいだろう

# テーブルディスカッションの論点

## 2. 国内で初めての開発

201試験と202試験はいずれも米国で実施された。日本では未承認(日本人の治験での投与経験なし)。海外の情報を使って日本で開発するとき、どのようなアプローチがあるか？

- ヒストリカルコントロールとして海外試験(201試験, 202試験)を利用し, 併合解析を試みる。  
日本での臨床試験の設計も海外試験を踏襲する形として, 結果の解釈の際に併合する。
- 日本で臨床試験を実施し, 治療効果を確認する。  
ICH E5(ブリッジング試験)の考え方を参照し, 国内承認申請につなげることも一案か。
  - PKの確認(民族差の観点で)
  - 日本人でも, ジストロフィンの改善が臨床症状改善につながるのか, など
    - 日本人集団と全体集団の一貫のベースとなる(遺伝子など)が変わっていない点は前提条件になるだろう
- 日本で臨床試験を実施する場合, 単群で実施することで計画することを初案として出したい。
- 日本以外で承認されていない国も含めた, 国際共同治験を実施することも選択肢ではないか。
  - この場合も民族的な背景などの考慮も必要となるだろう。
- 日本で試験を実施するか? 試験を実施せずに他の情報を借用する?
- 日本人でも同様の薬効が得られる, が成り立たない時は, 恐らく海外試験と同様にヒストリカルコントロールを用いた評価などを実施する必要があるだろう。
- ブリッジング試験(日本人への投与実績, オープン試験でも)
  - わずかな例数で良いか?
  - 統計的な根拠よりも, 日本人への投与実績と安全性の最低限の確認
- 薬理的な比較(代謝酵素など)
- PMDAとの事前の議論が重要になる
- 作用機序が似ている他剤との比較, 他剤の不応例を対象とした試験

# テーブルディスカッションの論点

## 3. 外部対照での感度分析

患者レジストリ186名から、患者選択基準に該当する13名のデータを抽出して、比較している。外部対照を用いることによるバイアスに対応するため、どのような方法があるか？

- 条件を付けて抽出するだけでなく、PSマッチングなどで類似した患者を抽出し、マッチングの結果どの程度類似している患者集団になったのか示すことも必要
- ①適格基準が同じ集団を抽出する。②背景を揃える、③エンドポイントの評価方法(データの取得方法、頻度など)が同じか、確認する
  - ③評価方法が異なる場合は...？選択肢から外れるのか？
    - 臨床的背景も含めた評価があるかもしれない。  
(多少の違いが見られた場合、それは同等と考えられるのか、など。感度分析により評価)
    - これから実施する試験の評価、であれば過去の試験にあわせることも選択肢となるのでは。
- 患者背景を揃える観点では、統計解析のなかで共変量の調整を実施することも必要(Crudeと調整済みの結果を確認する)。
  - レジストリデータであれば共変量のデータは含まれているはずである。
- ハイブリッドコントロールを使うことも一案では。
  - 試験の中でも対照群の患者のデータをとれていれば(今回の試験で利用できるかは要確認。)
- この試験でマッチングは適切なのか(例数の観点)...？  
ベースライン値と年齢のみ調整した解析までが現実的なのではないか。

# テーブルディスカッションの論点

## 3. 外部対照での感度分析

患者レジストリ186名から、患者選択基準に該当する13名のデータを抽出して、比較している。外部対照を用いることによるバイアスに対応するため、どのような方法があるか？

- 外部対照を用いることによるバイアスは様々考えられるが、それぞれのバイアスに対処するための感度分析を検討し、種々対応する必要がある
    - PSマッチングで調整できるバイアス？
    - 適格基準が同じ集団という抽出条件で調整できるバイアス？
    - 評価方法の違いを評価することで調整できるバイアス？
  - target trial emulation的な考え方もできそう
  - マッチング条件を1つずつ緩和して影響を確認する、別のレジストリを使ってみる、マッチングの方法に関する感度分析、共変量の選び方、Power priorのパラメータ設定、欠測値の補完、ベースライン時点を変えるなど
  - 共変量がそもそもデータベースに含まれているかも重要
  - 6分間歩行試験はそもそも適切か？（状況や気持ちで成績が異なる？）
    - 自律歩行ができなくなる等、客観的でハードな評価項目の方が良い。でも、レジストリには入っていないかな
- ...

# テーブルディスカッションの論点

## 4. Pocockの基準を再考

現代のrare diseaseでの治療開発にて, Pocockの基準は適合するか?  
基準に過不足はあるか?

- Best available therapyなどは時代と共に変化するので, 治療体系が変わるとそれより昔のデータは難しい。
- 「同じ組織」が難しい。他の要件を満たせば「同じ組織」は要らない? 同じ組織=同じ国と理解すれば現在でも有効では。
- 全て妥当で, 理想的には全て満たすべきだが, すべてを満たすことは難しい。ただし, 全てを完全に満たさなくても利用できるのでは。
- 基準を満たす程度によって情報借用の程度を調整する(dynamic borrowing)。
- 過去の臨床試験(RWDは想定されていない)が念頭にある基準かもしれない。
- 盲検でないヒストリカルデータの場合は, ハードなエンドポイントが望ましい。
- 評価タイミングが合わない場合もある。⇒ Visit windowを設定して対応する? モデルなどを用いて, ある時点のデータを予測する?
- サンプルサイズに影響を受けない, ベイズの階層モデルを用いる選択肢が用いられることも有るだろう
- Q: FDAの指摘で「治療の評価方法が新規試験と同じ」を満たしていない, とあるが, 満たしているのでは?
- 「新規試験と同じ適格基準を用いて最近実施された臨床試験である」
  - 寿命や標準治療の変化によるものと考えられる
    - 寿命や標準治療の変化(そもそも治療がない)が考えられない疾患領域であれば, 該当なしと考えられるのでは?



# テーブルディスカッションの論点

## 4. Pocockの基準を再考

現代のrare diseaseでの治療開発にて、Pocockの基準は適合するか？  
基準に過不足はあるか？

- 「新規試験とほぼ同じ組織で実施されている」は実現可能か？
  - ヒストリカルコントロールを作成した組織(病院施設)と、試験実施組織(病院施設)を同一とするのは難しいのではないか。
  - 「組織」の指すところを明確にした方が良い。
    - 現代では多施設・多地域試験の場合がある。
    - 地域？ Sponsor企業？ 実施施設？
      - 「地域」が異なっても、ある仮定を満たせば異なる地域のデータでも問題ないのでは？
      - 実施施設が試験によって異なるケースが考えられるのであれば、この条件はかなり厳しいものではないか。
    - 「地域」が異なっても、条件としては強くはない、と考えられるか？
      - 医療機関や地域によって、医療環境が異ならないような疾患(そもそも治療選択肢がない、などの条件)のであれば成り立つかもしれない。
  - 国、学会で希少疾患DBを作成し、管理されていれば、この問題はクリアされるだろう。
- True Endpointで評価できていれば、FDAが問題と挙げた点は問題と考えられたのか。
  - 実施組織のバラツキや治療の評価方法に関する問題は、結論には直結しないだろう
  - 画像評価において、中央判定することができれば、実施組織のバラつきは問題なくなるかもしれない
  - ヒストリカルコントロールを事後に決めた点については、評価項目がTrue Endpointでも問題だろう。

# (再掲) Pocockの基準 (1976)

---

## 対照群の一部に利用可能な既存のヒストリカルコントロールの条件

1. 詳細に定義された、新規試験の対照治療群と同じ治療を受けている。
2. 新規試験と同じ適格基準を用いて最近実施された臨床試験である。
3. 治療の評価方法が新規試験と同じである。
4. 重要な患者背景の分布が新規試験と同様である。
5. 新規試験とほぼ同じ組織で実施されている。
6. 新規試験との結果の違いを生むと予想される他の要因が存在しない。例えば、新規試験において症例登録スピードが予想よりも上がらないとき、適格基準をみたすぎりぎりの症例が登録されることも起こりうる。そういった場合、新規試験と既存試験で共通の適格基準を用いたとしても登録患者の背景が二つの試験間で異なってくる可能性がある。

# 参考文献

---

- 日本製薬工業協会 (2022). Rare diseaseの治療効果の推測法.  
[https://www.jpma.or.jp/information/evaluation/results/allotment/jtrngf00000085m-att/ds\\_202212\\_rare.pdf](https://www.jpma.or.jp/information/evaluation/results/allotment/jtrngf00000085m-att/ds_202212_rare.pdf)
- 日本製薬工業協会 (2023). 「Rare diseaseの臨床開発における治療効果推測の現状と課題」シンポジウム  
[https://www.jpma.or.jp/information/evaluation/symposium/2023\\_02\\_20.html](https://www.jpma.or.jp/information/evaluation/symposium/2023_02_20.html)
- 戸田達史. “2. 筋ジストロフィーの分子機構と治療戦略.” 日本内科学会雑誌 105.9 (2016): 1578-1587.
- Pocock, Stuart J. "The combination of randomized and historical controls in clinical trials." Journal of chronic diseases 29.3 (1976): 175-188.