【日本製薬工業協会シンポジウム】
Rare diseaseの臨床開発における治療効果推測の現状と課題

セッション1：Rare diseaseの治療効果の推測法
- ヒストリカル・コントロールの利用及びその実例 -

2023年2月20日

医薬品評価委員会　データサイエンス部会　KT3

〇ユーシービージャパン株式会社　五十川 直樹
〇日本新薬株式会社　森本 賢策
キッセイ薬品工業株式会社　岡村 正太

Disclaimer

製薬協

- 本演題の内容は、公開されている情報について、本タスクフォースにて集約・考察したものです。タスクフォースメンバーが所属する企業活動とは無関係であり、また企業の見解を示すものではありません。

- 演者は製薬会社の社員でありますが、本演題は本タスクフォースの立場で発表しています。

Agenda

製薬協

- 主なヒストリカルコントロールのリソース
- 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
- ヒストリカルコントロールを利用する解析
  - 階層モデル
  - Power prior
  - マッチング
- ヒストリカルコントロールの解析手法適用時の留意点
- 事例紹介
  - セルリポナーゼ アルファ
  - eteplirsen とビルトラルセン

主なヒストリカルコントロールのリソース（Ghadessi et al. (2020)[1]）

- Real World Data（RWD）
  - RWDの種類
    - 主なRWD：カルテ（Medical chart）、患者レジストリ（Patient registry）、自然歴研究（Natural history study）
    - その他のRWD：適応外使用の公表データ、様々なソースから定期的に収集された患者の健康状態などのデータ
  - RWDを利用することで，通常の臨床試験では得られないような，より幅広い患者集団での情報を得られる可能性がある
  - RWDは承認申請資料等への二次利用を前提とせずに設計・運用されていることも多く，品質や適切性等の観点から承認申請等に利用可能か検討する必要がある
- 完了した臨床試験
  - 同じ作用機序または同じ薬剤の完了した臨床試験の対照（プラセボまたは標準治療）群のデータ

- データの構造や品質はリソースによって様々であり、リソースの種類に応じた様々なバイアスや懸念点がある
- 臨床試験で利用する際にはこれらの点を十分に理解する必要がある

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

- ヒストリカルコントロールの利用可能性はいくつかの文書で言及されている
  - ICH E10ガイドライン（2001）[2]
  - Rare Disease: Natural History Studies for Drug Development (FDA (2019))[3]
  - Framework for FDA‘s Real-World Evidence Program（FDA (2018)）[4]
- ヒストリカルコントロールを単群試験の外部対照群としたり，事前に定義したアルゴリズムやマッチング等の統計手法を用いて，二重盲検試験の対照群と併合することが考えられる[5]
- ヒストリカルコントロールはランダム化や盲検化がされた試験の同時対照のデータではないため，以下のような潜在的バイアスの存在が懸念される
  - 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
- 潜在的バイアスを低減させるために以下のような方策が考えられ、試験実施前に決定しておく
  - 利用するヒストリカルコントロールの基準を設ける
  - 解析手法による対処

[5] Chen et al (2021)

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

製薬協

- **Pocockの基準[6, 7]**
  - **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)[6]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**

[6] Pocock (1976), [7] 武田ら（2015）

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の基準の条件を全て満たすヒストリカルコントロールデータが存在することは現実的にほとんどない
- Pocockの基準の条件が一部満たされていない状況等でも，バイアスを低減させながらヒストリカルデータの情報を利用する方法として，以下が考えられる
  - 階層モデル、Power prior
    - ヒストリカルデータと新規試験データの類似度に応じてヒストリカルデータから借用する情報量の大きさを制御する
  - 傾向スコア
    - ヒストリカルデータと新規試験データ患者の背景因子（ベースライン共変量）の不均一性を調整する

ヒストリカルコントロールを利用する解析

➢ 通常の階層モデル[8, 9]

- 階層モデルによるアプローチでは，新規試験の対照群のパラメータとヒストリカルコントロールのパラメータを同じ分布からのランダムサンプリングと仮定する。
- ヒストリカルコントロールデータの利用度は試験間のばらつき（試験間分散）の大きさによる。試験間のばらつきがない場合はヒストリカルコントロールデータをすべて利用し，試験間のばらつきが大きくなるに従ってヒストリカルコントロールデータの利用度は小さくなる。
- Meta-analytic prior (MAP)は，ヒストリカルコントロールデータに基づいて予測分布を推定し，この予測分布を新規試験の対照群の事前分布として利用して，パラメータの推定を行う。

超事前分布

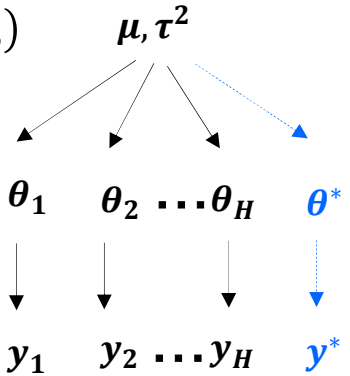
$$\mu \sim N(\mu_\phi, \tau_\phi^2), \tau^2 \sim IGa(a_\phi, b_\phi)$$

初期事前分布

$$\theta_1, \ldots, \theta_H, \theta^* \sim N(\mu, \tau^2)$$

サンプリングモデル

$$Y_h|\theta_h\sigma_h \sim N(\theta_h, \sigma_h^2)$$



θ^* の予測分布+ → MAP prior ($p_{MAP}(\theta^*)$)

$$\theta^*|Y_1, \ldots, Y_H, \tau \sim N\left(\frac{\sum w_h Y_h}{\sum w_h}, \frac{1}{\sum w_h} + \tau^2\right)$$

$$w_h = \left(\sigma_h^2 + \tau^2\right)^{-1}$$

+) 試験内・試験間の分散既知，μは一様分布を仮定

[8] Neuenschwander et al (2010), [9] Gsteiger et al. (2013)

ヒストリカルコントロールを利用する解析

- 階層モデル（Robust MAP prior）[10]
  - 現実には，新規試験の対照群データとヒストリカルコントロールデータが類似しておらず，これらのパラメータの交換可能性が成り立たない可能性も考えられる。このような場合を考慮した頑健な方法として，Robust MAP priorが提案されている。
  - MAP priorを$p_{MAP}(\theta^*)$，曖昧な事前分布（Vague prior）を$p_{vague}(\theta^*)$とすると，Robust MAP priorは以下の式で表される。ここで，wは新規試験の対照群データとヒストリカルコントロールデータが類似していない事前確率と解釈することができ，ヒストリカルコントロールデータの借用度に対する懐疑性の程度を表している。

$$p_{R-MAP}(\theta^*) = (1-w)p_{MAP}(\theta^*) + wp_{vague}(\theta^*)$$

超事前分布

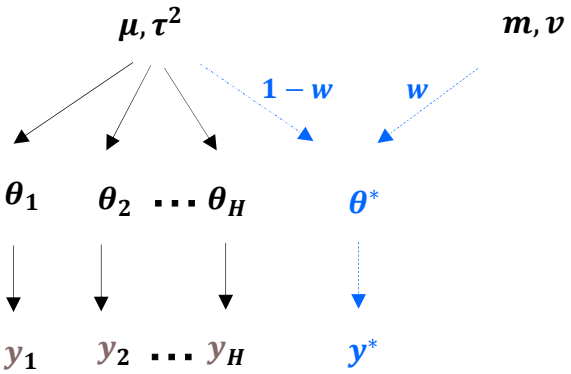
$$\mu \sim N(\mu_\phi, \tau_\phi^2), \tau^2 \sim IGa(a_\phi, b_\phi)$$

初期事前分布

$$\theta_1, \ldots, \theta_H, \theta^* \sim N(\mu, \tau^2)$$

サンプリングモデル

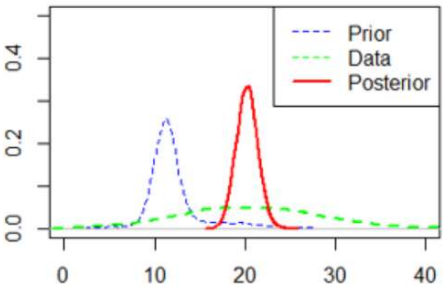
$$Y_h|\theta_h\sigma_h \sim N(\theta_h, \sigma_h^2)$$



曖昧な事前分布

$$\theta^* \sim N(m, v)$$

Prior conflict



[10] Schmidli et al (2014)

ヒストリカルコントロールを利用する解析

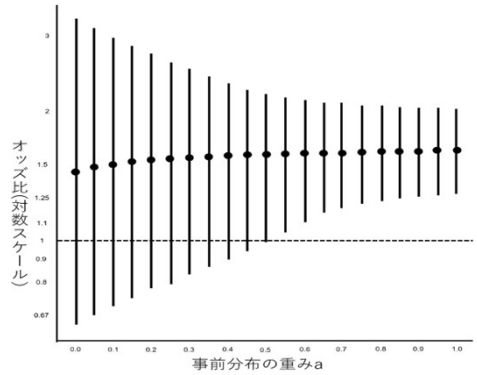
製薬協

- その他の階層モデル（Bayesian semiparametric MAP）[11]
  - Bayesian semiparametric MAPでは，MAP priorでの試験間変動をディリクレ過程混合モデル（Dirichlet process mixture model）でノンパラメトリックにモデル化することで，新規試験の対照群データとヒストリカルコントロールデータの類似性だけでなく，ヒストリカルコントロールデータ間の類似性も考慮して，ヒストリカルコントロールデータの借用度を調整する。

  - 希少疾病以外の活用事例
    - PoC試験（Secukinumab，クローン病）：6 つの既存の臨床試験から得られたプラセボ群 671 例の結果から，20 例分の情報を持つようにプラセボ群の事前分布を設定した[12]。
      - ✓ MAP prior: $\theta_P|Y_{P1},\ldots,Y_{P6} \sim \mathrm{N}[-49, \sigma^2/20]$
    - 小児試験（ベリムマブ，SLE）：FDAからの照会事項に基づき、過去の成人データを借用し，小児患者の有効性を評価した[13]。
      - ✓ Robust MAP prior: $\delta_p \sim (1-a) \times N(0, 8.27) + a \times N(0.48, 0.015)$

        Skeptical prior　　成人データ

        a：事前分布（成人データ）の重み



[11] Hupf et al (2021), [12] Hueber et al (2012), [13] FDA (2019)

ヒストリカルコントロールを利用する解析

- Power prior[14]
  - Power priorは初期事前分布と情報量を割り引かれたヒストリカルコントロールデータの尤度から得られた事後分布を指し，新規試験のデータを解析する際にヒストリカルコントロールデータの情報を有した事前分布として用いられる。
  - 関心のあるパラメータの初期事前分布と1つの過去の試験からのヒストリカルコントロールデータの尤度をべき乗パラメータ（$a_0 \in [0, 1]$）でべき乗した項の積として定義される。
    - 詳細は次スライド
  - 希少疾病以外に提案されている活用場面
    - Phase 2-3[15]
    - Bridging studies between Caucasians and Asians in Ph1[16]

[14] Ibrahim et al (2000), [15] Rietbergen et al (2011), [16] Ollier et al (2020)

ヒストリカルコントロールを利用する解析

製薬協

➢ Power prior[17–20]

Power prior

べき乗パラメータが分布に従う

尤度原理を満たさない

Joint power prior

$$p(\theta, a_0|D_0) \propto L(\theta|D_0)^{a_0} p_0(\theta) p(a_0)$$

複数のデータ

Multiple joint power prior

$$p(\theta, a_0|D_{01}, \ldots, D_{0H}) \propto \left\{ \prod_{h=1}^{H} L(\theta|D_{0h})^{a_{0h}} p(a_{0h}) \right\} p_0(\theta)$$

尤度原理を満たす

Modified power prior

$$p(\theta, a_0|D_0) \propto \frac{L(\theta|D_0)^{a_0} p_0(\theta)}{\int L(\theta|D_0)^{a_0} p_0(\theta) d\theta} p(a_0)$$

複数のデータ

Multiple modified power prior

$$p(\theta, a_0|D_{01}, \ldots, D_{0H}) \propto \frac{1}{C(a_1, \ldots, a_H)} \left\{ \prod_{h=1}^{H} L(\theta|D_{0h})^{a_{0h}} p(a_{0h}) \right\} p_0(\theta)$$

Commensurate power prior

データの共通性を活用する

$$p(\theta, a_0, \tau|D_0) \propto \frac{L(\theta_0|D_0)^{a_0} p_0(\theta_0)}{\int L(\theta_0|D_0)^{a_0} p_0(\theta_0) d\theta_0} p_0(\theta|\theta_0, \tau)\, p(a_0|\tau) p(\tau)$$

Dependent modified power prior

ヒストリカルコントロールデータ間の類似性を考慮する

$$p(\theta, a_{01}, \ldots, a_{0H}, \mu, \sigma^2|D_{01}, \ldots, D_{0H}) \propto \frac{1}{C(a_1, \ldots, a_H)} \left\{ \prod_{h=1}^{H} L(\theta|D_{0h})^{a_{0h}} p(a_{0h}|\mu, \sigma^2) \right\} p_0(\theta) p(\mu) p(\sigma^2)$$

Conditional power prior

べき乗パラメータが分布に従わない

$$p(\theta|D_0, a_0) \propto L(\theta|D_0)^{a_0} p_0(\theta)$$

複数のデータ

Multiple conditional power prior

[17] Duan et al (2006), [18] Neuenschwander et al (2009), [19] Hobbs et al (2011), [20] Banbeta et al (2019)

ヒストリカルコントロールを利用する解析

製薬協

- マッチング（Matching）
  - 単純なマッチング
    - 新規試験データの患者について，調整したい背景因子に完全に一致するような外部試験データの患者を選ぶ方法を単純なマッチングと呼ぶ。しかしながら，単純なマッチングで考慮できる因子数は限られるため，傾向スコア等を使ったマッチングがよく用いられる。
  - 傾向スコア（Propensity Score）[21]
    - 複数の交絡因子を調整し，因果効果を推定するための方法
    - ここでは，単群試験または比較試験の対照群に対して，外部試験データを活用して，新規試験データを補う目的
    - Z=1が新規試験データ，Z=0が外部試験データを表し，Xをベースライン共変量とするとき、傾向スコアは以下で表される。

$$e(\boldsymbol{X}) = \Pr[Z = 1|\boldsymbol{X}]$$

  - 傾向スコアに基づくMAP priorアプローチ[22]
    - 推定した傾向スコアをいくつかの層に分割し，層内の新規試験データと外部試験データの傾向スコアの類似性を評価することで，層間の不均一性を考慮した上で，層ごとに治療結果を推定し，最後に併合することで全体の治療結果を求める方法
    - 傾向スコアに基づくPower priorアプローチ[23]も提案されている。

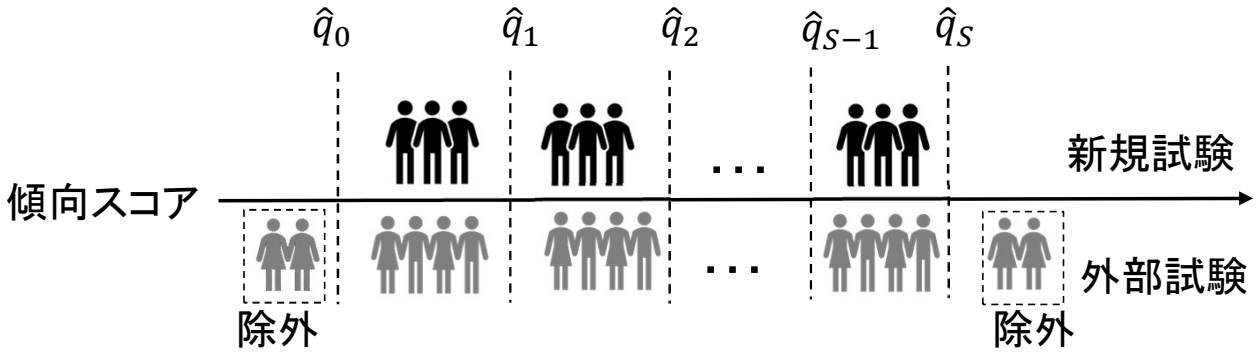
[21] Rosenbaum et al (1983), [22] Liu et al (2021), [23] Wang et al (2019)

ヒストリカルコントロールを利用する解析

- マッチング（Matching）
  - 傾向スコアに基づくMAP priorアプローチ

[Step 1] 傾向スコアの推定，及び切り取り（Trimming）
各被験者iの傾向スコア$\hat{e}(X_i)$を求める。$E_1 = \{\hat{e}(X_i): Z_i = 1\}$を新規試験の患者から推定した傾向スコアの集合。
E_1の範囲外の傾向スコアをもつ外部試験データの患者を除外する。

[Step 2] 層別化
Step 1で得られた患者集団に対して，傾向スコアによる層別化を行い，S個の層に分ける。
sでs番目の層を示し（$s = 1, \ldots, S$），s番目の層に該当する傾向スコアの患者の共変量の部分集団を$X_s = \{X: \hat{e}(X) \in (\hat{q}_{s-1}, \hat{q}_s]\}$，治療群zの治療結果の部分集合を $Y_{s,z} = \{y_i: X_i \in X_s, Z_i = z\}$で表すとする。



ヒストリカルコントロールを利用する解析

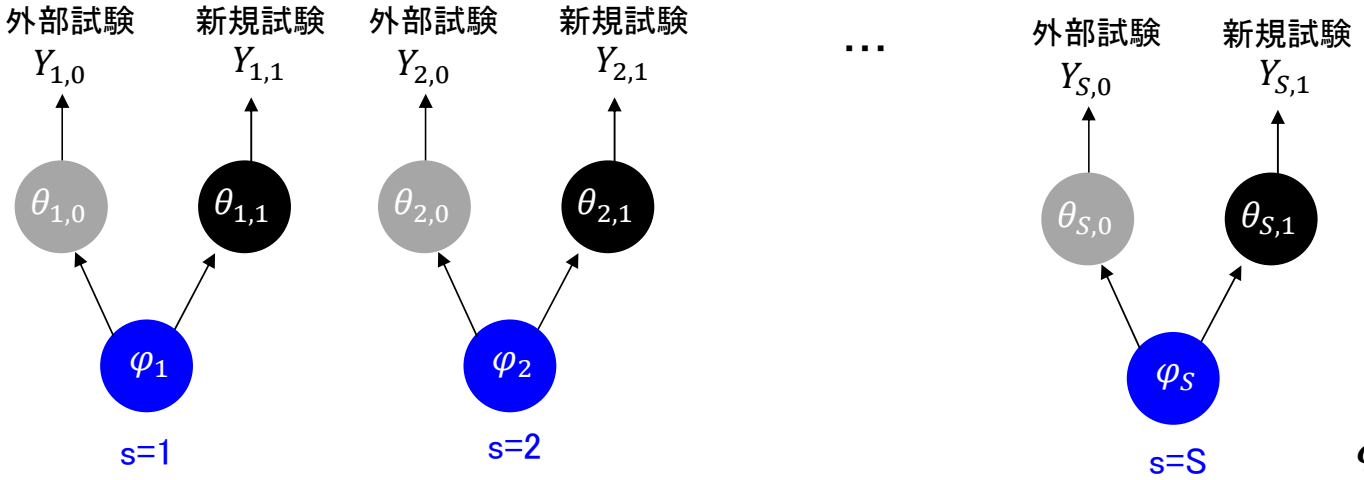
- マッチング（Matching）
  - 傾向スコアに基づくMAP priorアプローチ

[Step 3] 層特有のMAP prior

s番目の層の$Y_{s,Z} = \{y_i : X_i \in X_s, Z_i = z\}$ を連続量の結果としたとき，ベイズ流階層モデルは以下の通りである。

・$Y_{s,z}|\theta_{s,z} \sim \mathrm{N}\left[\theta_{s,z}, s_{s,z}^2\right],\ s = 1, \ldots, S, z = 0, 1$

・$\theta_{s,z}|\mu_s, \tau_s \sim \mathrm{N}\left[\mu_s, \tau_s^2\right],\ s = 1, \ldots, S, z = 0, 1$



φ_s：層sのハイパー・パラメータ

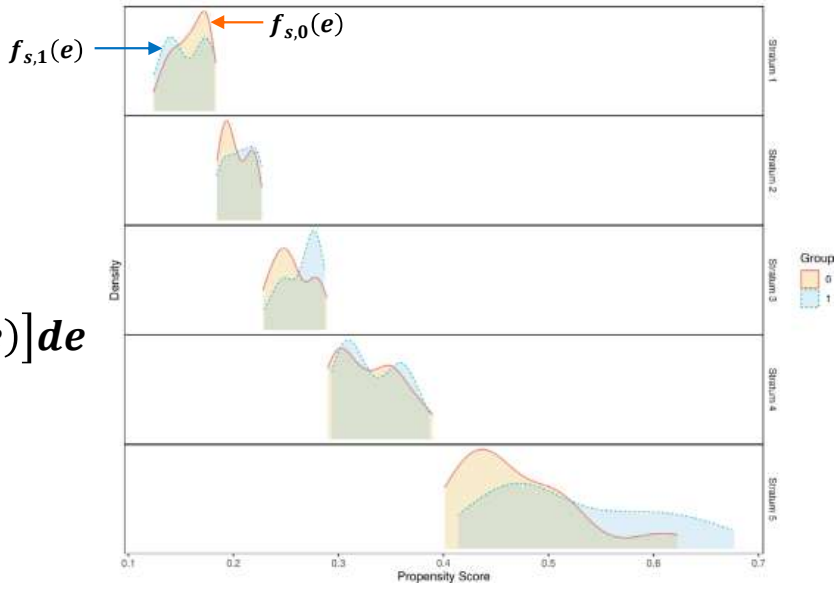
新規試験データの全体の治療結果θ_1（$p_{PS-\mathrm{MAP}}(\theta_1) = p(\theta_1|X_0, X_1, Y_0)$）は層特有のMAP priorの治療結果の重み付き平均となる。

$$\theta_1 = \sum_{s=1}^{S} \frac{n_{s,1}}{N_1} \theta_{s,1}, N_1 = \sum_{s=1}^{S} n_{s,1}$$

ヒストリカルコントロールを利用する解析

- マッチング（Matching）
  - 傾向スコアに基づくMAP priorアプローチ

[Step 4] 傾向スコアに基づくMAP prior（PS-MAP prior）
パラメータμ_sの事前分布：曖昧な事前分布
パラメータτ_sの事前分布：$\tau_s \sim \text{half normal}[k_s \times t]$
・k_sは各層の類似性を反映，tは層特有の分散をスケール化。
・$k_s = \frac{r^{ref}}{r_s}$，$r^{ref}$としてRの中央値等，$r_s = \int_0^1 \min\left[f_{s,0}(e), f_{s,1}(e)\right] de$
・Effective sample sizeに基づき更新される。

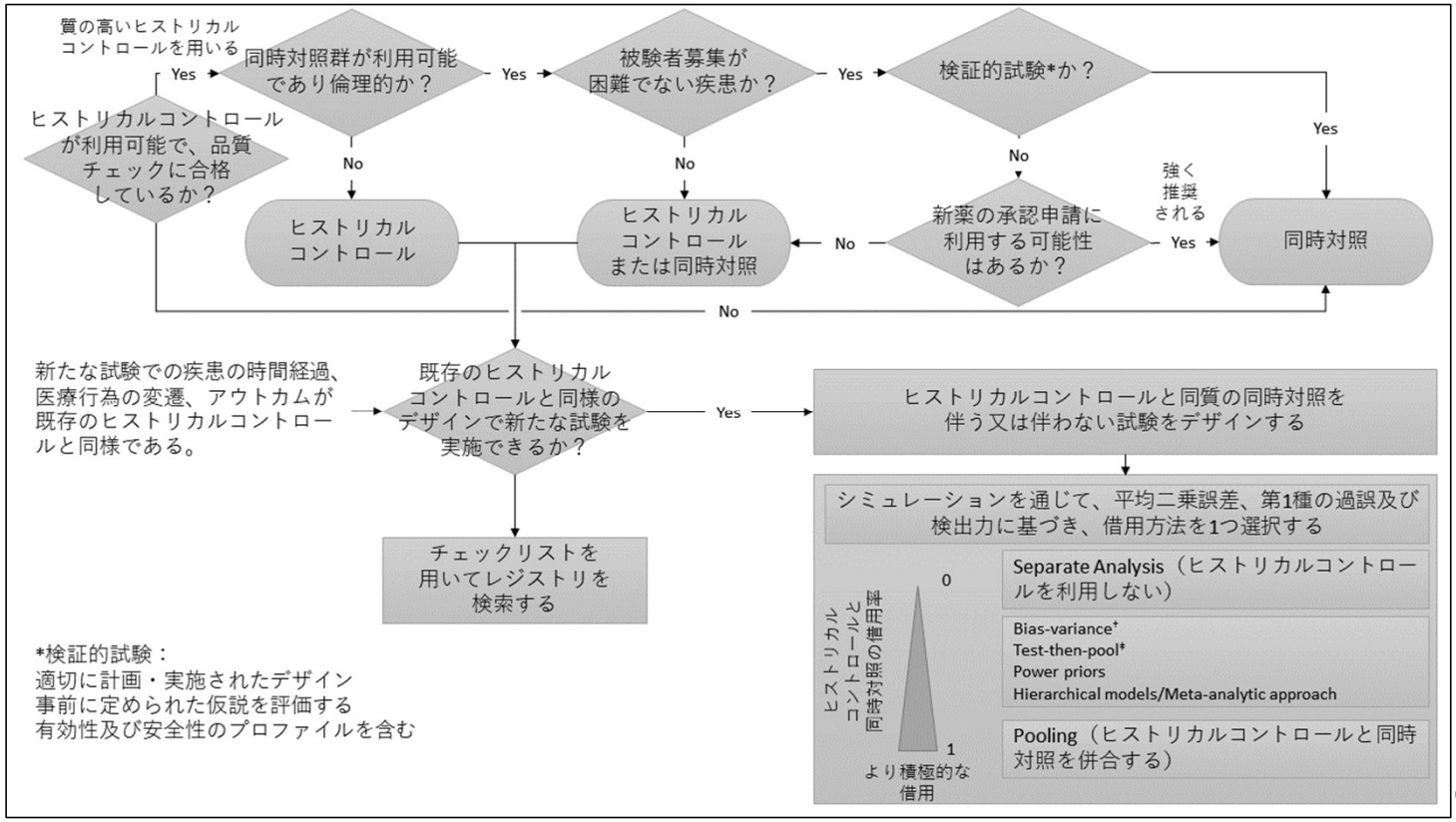


[Step 5] 解析結果
新規試験の結果を$p(Y_1|\theta_1)$（Y_1は新規試験データの目的変数）でモデル化し，Step 3-4で求めたPS-MAP priorを用いて，新規試験データの関心のあるパラメータθ_1の事後分布は以下で求められる。

$$p(\theta_1|Y_1) \propto p(Y_1|\theta_1)\, p_{PS-\text{MAP}}(\theta_1)$$

ヒストリカルコントロールの解析手法適用時の留意点

- 臨床試験でヒストリカルコントロールを利用するためのDecision Making Diagram[1]
  - 同時対照の設定や解析手法の検討に役立つ



[1] Ghadessi et al (2020)改編

ヒストリカルコントロールの解析手法適用時の留意点

製薬協

- シミュレーションを実施して解析手法の特性を把握する
  - 解析手法が最も威力を発揮する範囲(平均二乗誤差が少なく、第1種の過誤確率が低く、検出力が高い)を手法間で比較する
  - 同時対照群とヒストリカルコントロールデータが類似している場合
    - 第1種の過誤確率が減少し、検出力が向上する
  - 同時対照群とヒストリカルコントロールデータが類似していない場合
    - 第1種の過誤確率が増大することや、検出力が低下することもある
  - 既知の外部情報や潜在的バイアスのために、第1種の過誤確率が増大することは、規制当局として懸念がある[5]
    - 解析手法の特性の評価は、ヒストリカルコントロールを利用するための設定が適切であるかを判断する際の重要な材料となる

[5] Chen et al (2021)

ヒストリカルコントロールの解析手法適用時の留意点

製薬協

- 感度分析の実施
  - 主解析の結果の安定性を評価するための感度分析の際に考慮する事項として以下が考えられる[1]
    - 選択バイアスの影響
    - 主要な結果に対する欠測データの影響
    - ヒストリカルコントロールと新規試験データのマッチング方法
    - 異なる事前分布や方法を用いた場合のヒストリカルコントロールの借用度
    - 重み付け係数の変更
    - 複数のヒストリカルコントロールを用いる場合に，一部のヒストリカルコントロールを除いたときの影響
    - Pocockの基準を満たしていないヒストリカルコントロールの影響
    - （ヒストリカルデータがある場合の）ヒストリカルデータ間の異質性，及びヒストリカルコントロールデータと同時対照データの異質性

[1] Ghadessi et al (2020)

ヒストリカルコントロールの解析手法適用時の留意点

製薬協

- 得られている情報や新規試験の目的に応じて解析手法を選択する
  - ヒストリカルコントロールとして活用できる過去試験が少数：Power prior
  - 過去試験が複数存在し，試験間に相関があると想定：階層モデル
  - ヒストリカルコントロールの患者の背景因子の情報が利用可能：傾向スコア

- ヒストリカルコントロールの活用する際には、以下を行うことが重要
  - 解析手法の選択前にヒストリカルコントロールが利用可能かの判断
  - 新規試験開始前に選択する解析手法についての動作特性の把握
  - 新規試験データが得られた後の解析手法の適切性を評価する感度分析

事例紹介：ヒストリカルコントロール

ヒストリカルコントロールを利用した事例

薬剤名	対象疾患	事例としての特徴
セルリポナーゼアルファ	セロイドリポフスチン症2型	試験群とヒストリカルコントロール群の評価指標，患者背景の違いについて指摘を受けた事例
nifurtimox	アメリカトリパノソーマ症（シャーガス病）	主要評価項目の臨床的意義について審査過程で懸念が示され，ヒストリカルコントロールとの比較に基づく薬効評価が困難となった事例。
eteplirsen	デュシェンヌ型筋ジストロフィー	ヒストリカルコントロール群との比較を試みたものの，規制当局からは否定的な見解が得られた。
ビルトラルセン	デュシェンヌ型筋ジストロフィー	試験で収集する評価項目をレジストリの収集項目に合わせることによって薬効評価の比較可能性を高めた事例。

ヒストリカルコントロールを利用した事例

薬剤名	対象疾患	事例としての特徴
セルリポナーゼアルファ	セロイドリポフスチン症2 型	試験群とヒストリカルコントロール群の評価指標，患者背景の違いについて指摘を受けた事例
nifurtimox	アメリカトリパノソーマ症（シャーガス病）	主要評価項目の臨床的意義について審査過程で懸念が示され，ヒストリカルコントロールとの比較に基づく薬効評価が困難となった事例。
eteplirsen	デュシェンヌ型筋ジストロフィー	ヒストリカルコントロール群との比較を試みたものの，規制当局からは否定的な見解が得られた。
ビルトラルセン	デュシェンヌ型筋ジストロフィー	試験で収集する評価項目をレジストリの収集項目に合わせることによって薬効評価の比較可能性を高めた事例。

事例1 ｜セルリポナーゼ アルファ（ブリニューラ®）

製薬協

セロイドリポフスチン症2型（CLN2）について[26]

ライソゾーム内の TPP1※ の欠損を特徴

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中央値は 7.8年

CLN2の治療法は長らく対症療法、緩和療法のみ

＊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

[26] BioMarin Pharmaceutical Japan 株式会社. (2022).

事例1 | セルリポナーゼ アルファ(ブリニューラ®)

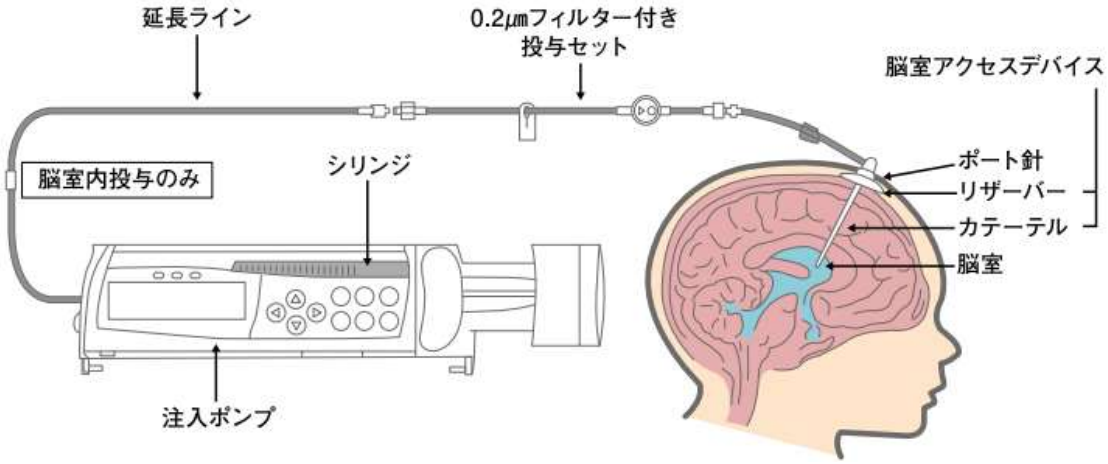
セルリポナーゼ　アルファについて

BioMarin社が開発

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 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[27]の植込み手術の実施



[27] BioMarin Pharmaceutical Japan 株式会社. (2021).

事例1 | セルリポナーゼ アルファ(ブリニューラ®)

製薬協

承認状況

2017年4月に米国で初めて承認

2017年5月に欧州で承認

2019年9月に日本で承認

事例1 | セルリポナーゼ アルファ(ブリニューラ®)

有効性評価に用いられた試験[28]

190-201試験(第I/II相)　実薬単群、非盲検、48週投与

190-202試験(第I/II相)　延長投与試験　(最長239週)

米国，ドイツ，イタリア，イギリスで実施

24例のCLN2患者が参加し，日本人患者はドイツにて1例参加

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

事例1 | セルリポナーゼ アルファ(ブリニューラ®)

製薬協

主要評価項目:CLN2評価尺度[29]

運動，及び言語それぞれ症状に応じて 0～3点のスコアを算出

合計スコアを算出し，最終的に 0点(重い障害)～6点(正常)のスコアとした

投与48週後における CLN2評価尺度がベースラインから 2点以上低下しない割合※

スコアの算出方法

オリジナルの CLN2評価尺度から評価をより明確化して改良した尺度を使用

※ ベースラインが1点の場合は0点にならない割合

[29] Steinfeld, Robert, et al. (2002)

事例1 | セルリポナーゼ アルファ(ブリニューラ®)

ヒストリカルコントロールの利用

DEM-CHILDデータベースに登録されているCLN2患者 69例

基準を全て満たす患者42例を 190-201/202試験と比較可能な患者として選定

- CLN2評価スコアが利用できる患者
- 190-201/202試験に参加しなかった患者
- 一卵性双生児を除いた患者
- 最初のCLN2評価から6か月以上経過した後に，1回以上評価された患者

スコアの算出方法

オリジナルのCLN2評価尺度を利用

スコアの算出方法の違い[30]

	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	理解不能または発語なし

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

FDAの審査から読み取れること：評価方法について

CLN2評価尺度に関するFDAからの指摘[28]

1. 評価尺度の違い

言語尺度で190-201/202試験の方のスコアが高くなる可能性

2. 評価方法の違い

<u>190-201/202試験</u>

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

<u>ヒストリカルコントロール</u>

保護者へのインタビューまたはカルテレビューによる後ろ向きな評価

前向きな評価のデータが一部混在

同じ被験者でも評価方法が異なる

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

FDAの審査から読み取れること：評価方法について

CLN2評価尺度に関するFDAからの指摘[28]

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

ヒストリカルコントロールでは不定期

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

FDAの審査から読み取れること：評価方法について

FDAは201/202試験における CLN2評価尺度の再評価を要求

- 評価時のビデオ映像を使用
  - 190-201/202試験の trainer
  - CLN2 developer
- 190-201/202試験におけるもとの結果との一致性を確認

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[28]

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

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

FDAの審査から読み取れること：比較可能性について

190-201/202試験とヒストリカルコントロールの患者背景の違い[28]

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

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

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

FDAの審査から読み取れること：比較可能性について

1:1のマッチング解析

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

必要に応じて性別など追加の項目を加えて、1対1の関係になるように患者選択

その結果，双方の試験から 17組の患者が選択

主要評価項目

CLN2評価尺度の運動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点にならない被験者の割合

CLN2運動尺度の結果[28]

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[31]

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

事例3，4｜eteplirsen とビルトラルセン

製薬協

DMD（デュシェンヌ型筋ジストロフィー）[32]

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

[32] 戸田達史. (2016)

事例3，4｜eteplirsen とビルトラルセン

2つの薬剤について[32]

eteplirsen とビルトラルセンはエクソン・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（以下、DMD）の症状悪化を緩和させる薬剤

[32] 戸田達史. (2016)

事例3，4｜eteplirsen とビルトラルセン

承認状況

eteplirsen[33]

2016年9月米国で承認

日本では未承認

ビルトラルセン[34]

2020年3月日本で承認

2020年8月米国で承認

[33] ミクスOnline. [34] 医薬品医療機器総合機構. (2020).

事例3，4｜eteplirsen とビルトラルセン

製薬協

共通点

DMDを対象としている点

主要評価項目にジストロフィンに関する代替指標を使用

ヒストリカルコントロールを利用して運動機能の評価を実施

ヒストリカルコントロールを利用した運動機能の評価に関して

eteplirsenではFDAから否定的なコメント

ビルトラルセンではPMDAから肯定的なコメント

→　一見すると評価が分かれたように見える

※ビルトラルセンの運動機能の評価に対する、FDAのコメントは公開されていない

有効性評価に用いられた試験（eteplirsen）[35]

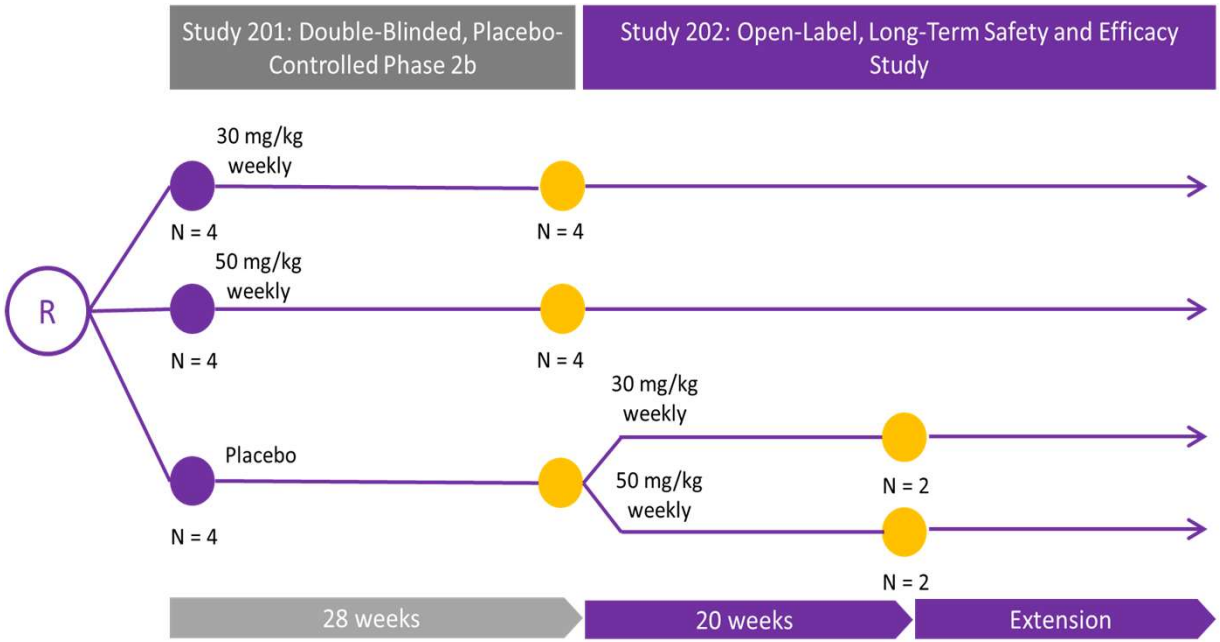
製薬協

201試験　ランダム化二重盲検プラセボ対照試験
米国の1施設で実施
12名のDMD患者が登録

202試験　非盲検の継続長期投与試験
201試験の終了後12名の被験者全員が移行

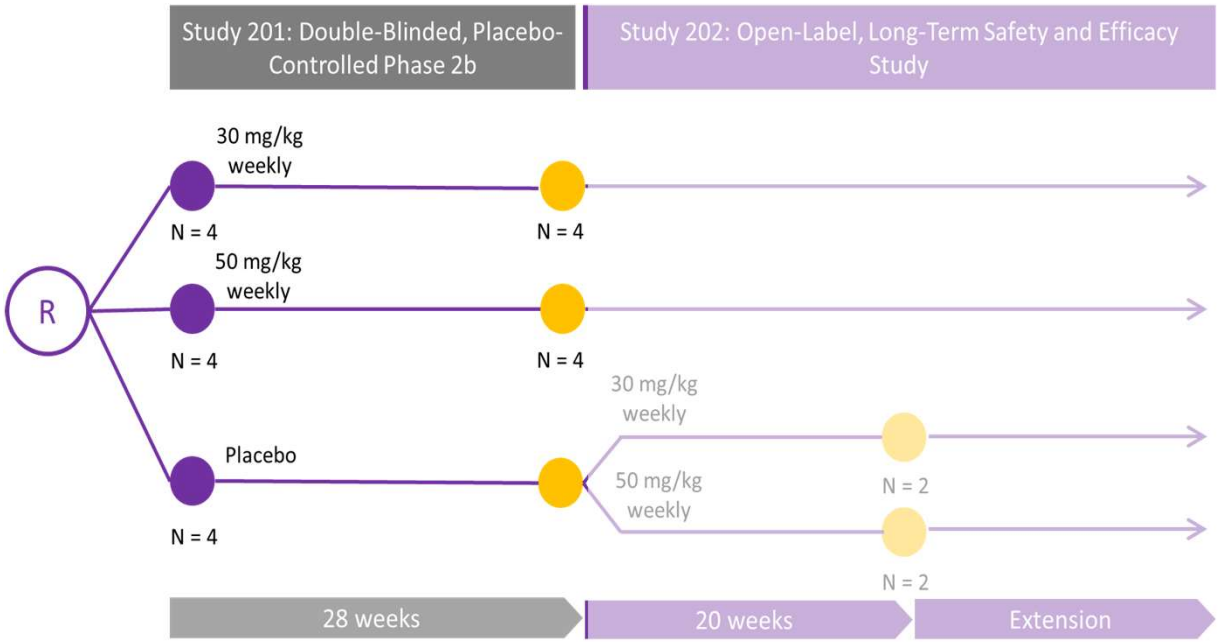
[35] Food and Drug Administration. (2016).

試験デザイン｜eteplirsenの試験（201及び202試験）[35]



[35] Food and Drug Administration. (2016).

試験デザイン｜eteplirsenの試験（201及び202試験）[35]



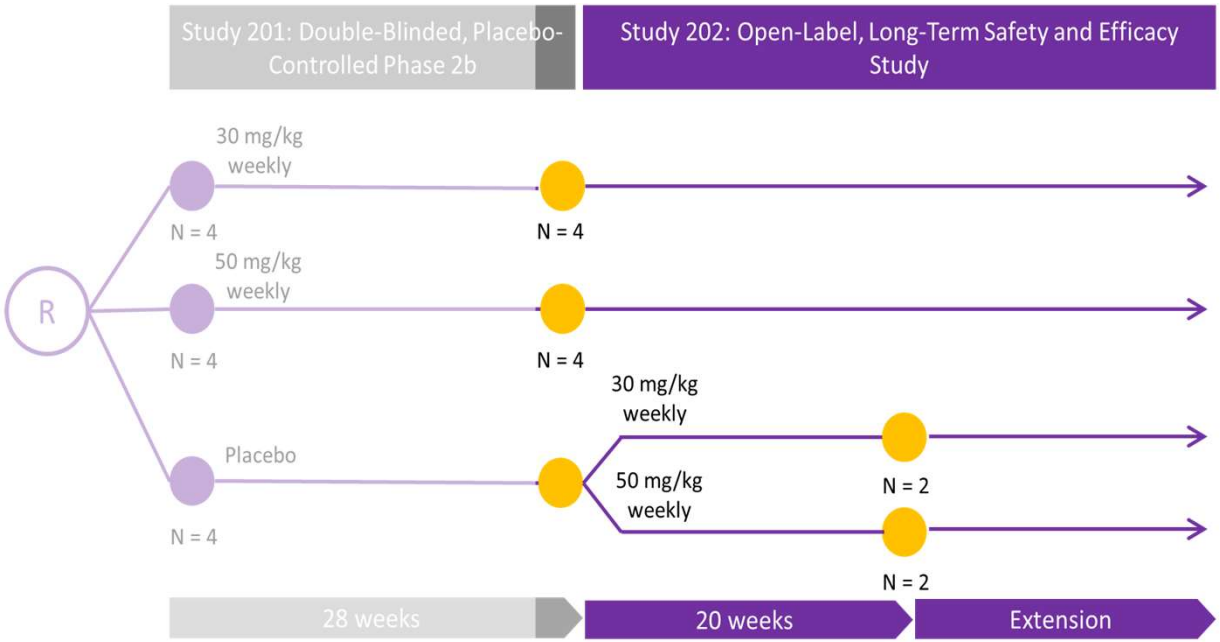
201試験

主要評価項目は筋肉生検組織中のジストロフィン陽性繊維の割合

20週時点においてプラセボ群と30mg/kg 群、50mg/kg 群との比較が行われた

[35] Food and Drug Administration. (2016).

試験デザイン｜eteplirsenの試験（201及び202試験）[35]



202試験　継続長期投与試験

主要評価項目は 6分間歩行試験

オープンラベル試験
全例 eteplirsenが投与

計48週時点の評価に疾患レジストリのデータをヒストリカルコントロールとして利用

[35] Food and Drug Administration. (2016).

主要評価項目の結果（eteplirsen）

筋肉生検組織中のジストロフィン陽性繊維の割合の結果[35]

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
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。

30 mg/kg群では22.95%の増加が認められた一方で、50 mg/kg群では0.79%の増加しか認められなかった

[35] Food and Drug Administration. (2016).

運動機能評価の結果（eteplirsen）

ヒストリカルコントロールを用いた6分間歩行試験の解析結果[35]

群		年齢	6分間歩行試験（ベースライン）	6分間歩行試験（36ヵ月後）**
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ヵ月後）に対して算出

** LS Meanの群間差（36ヵ月後）は141mであり，p値は0.009

eteplirsenが良好である結果が得られた

[35] Food and Drug Administration. (2016).

FDAからのコメント（eteplirsen）

「ヒストリカルコントロールを用いた非盲検の継続長期投与試験は，統計学的に解釈できない」という評価

指摘内容

1. 「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ICH E10[36]の条件にeteplirsenは該当しない
2. 6分間歩行試験はモチベーションによる影響を受け、客観的な指標ではない
3. Pocock の「治療の評価方法が新規試験と同じである」、「新規試験とほぼ同じ組織で実施されている」という基準を満たしていない
4. ヒストリカルコントロール群は事後的に特定
5. 対照群の選定や選択基準の設定は事前に計画されるべき

[36] 厚生労働省. (2001).

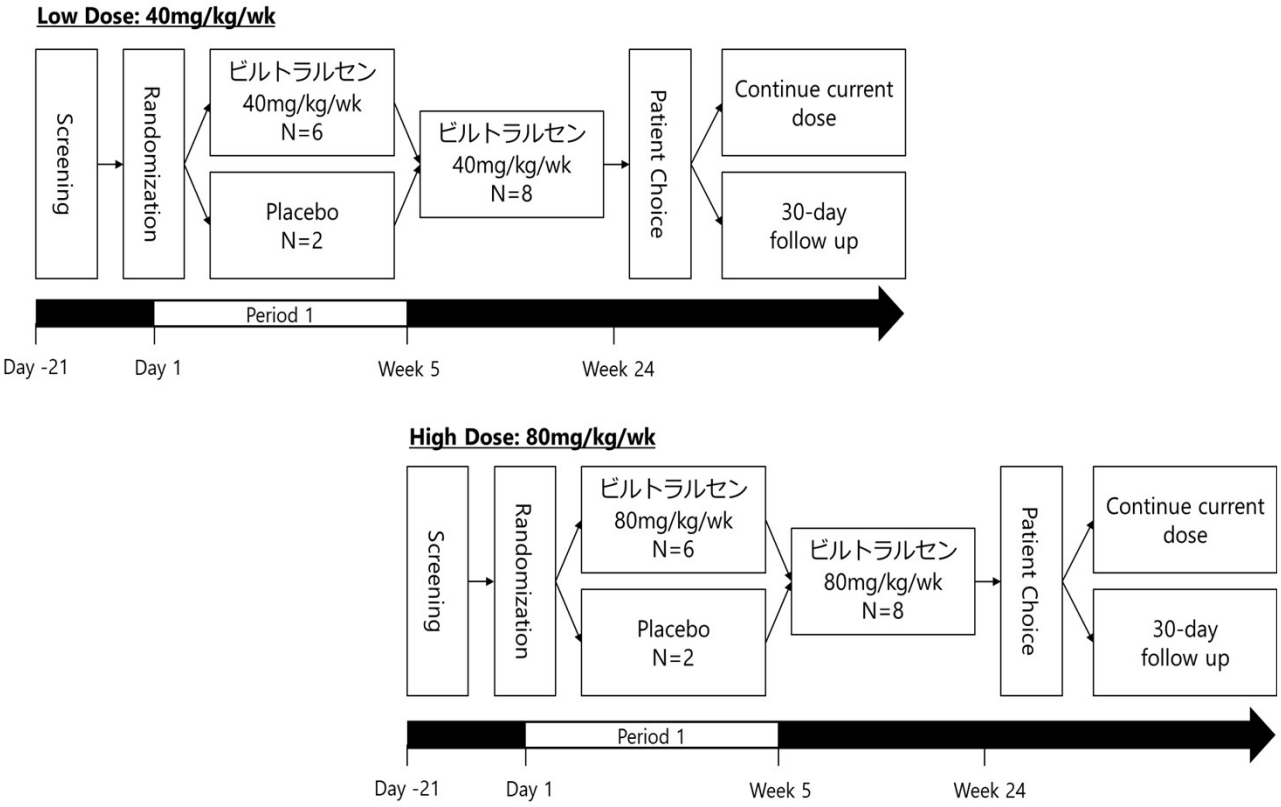
有効性評価に用いられた試験（ビルトラルセン）[34]

対照をおいて有効性評価を実施した試験

201試験　　ランダム化二重盲検プラセボ対照試験
米国及びカナダの6施設で実施
16名のDMD患者が登録

[34] 医薬品医療機器総合機構. (2020).

試験デザイン｜ビルトラルセンの試験

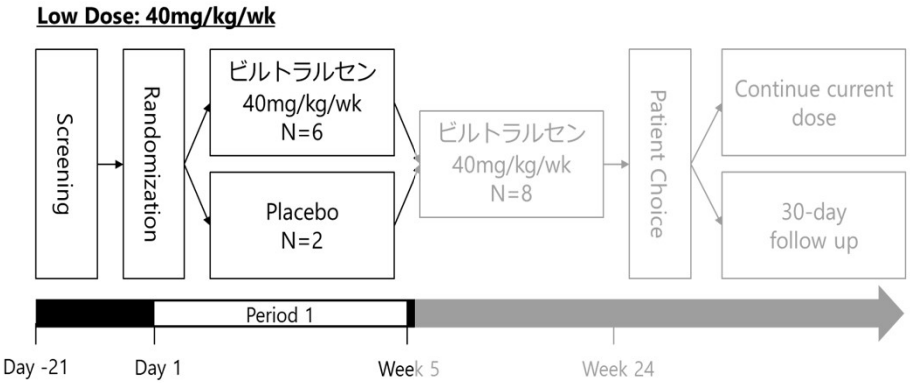


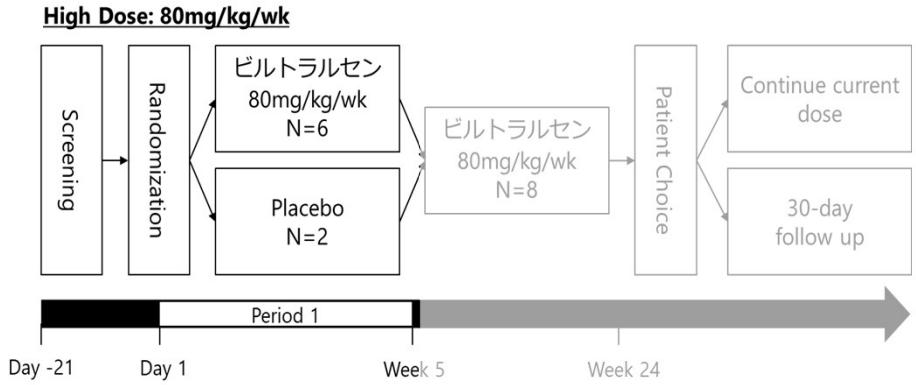
201試験　二重盲検試験

プラセボを対照においているが
投与後の有効性の評価は
Week 24で実施されるため
有効性の評価という観点では
事実上プラセボ対照ではない

主要評価項目はジストロフィン
タンパク発現

試験デザイン｜ビルトラルセンの試験



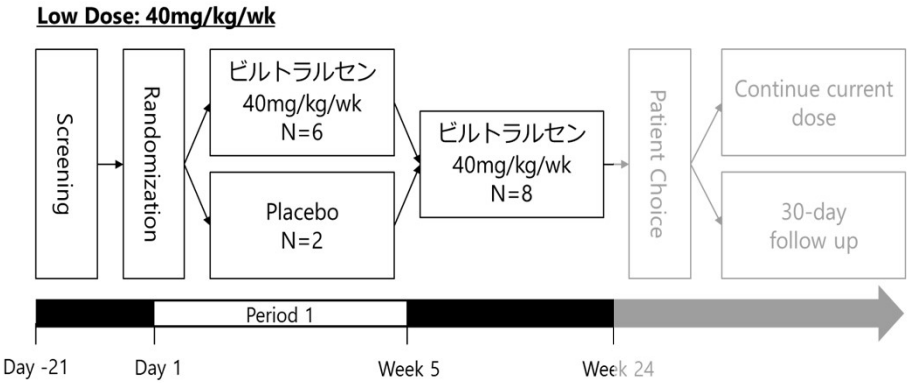


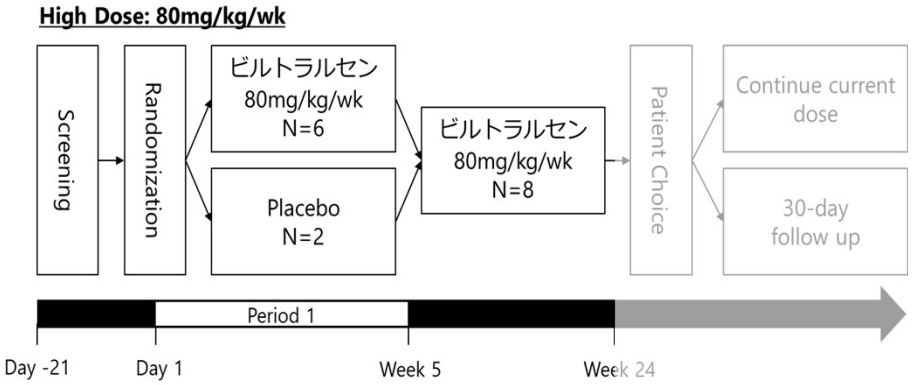
201試験　二重盲検試験

プラセボを対照においているが
投与後の有効性の評価は
Week 24で実施されるため
有効性の評価という観点では
事実上プラセボ対照ではない

主要評価項目はジストロフィン
タンパク発現

試験デザイン｜ビルトラルセンの試験



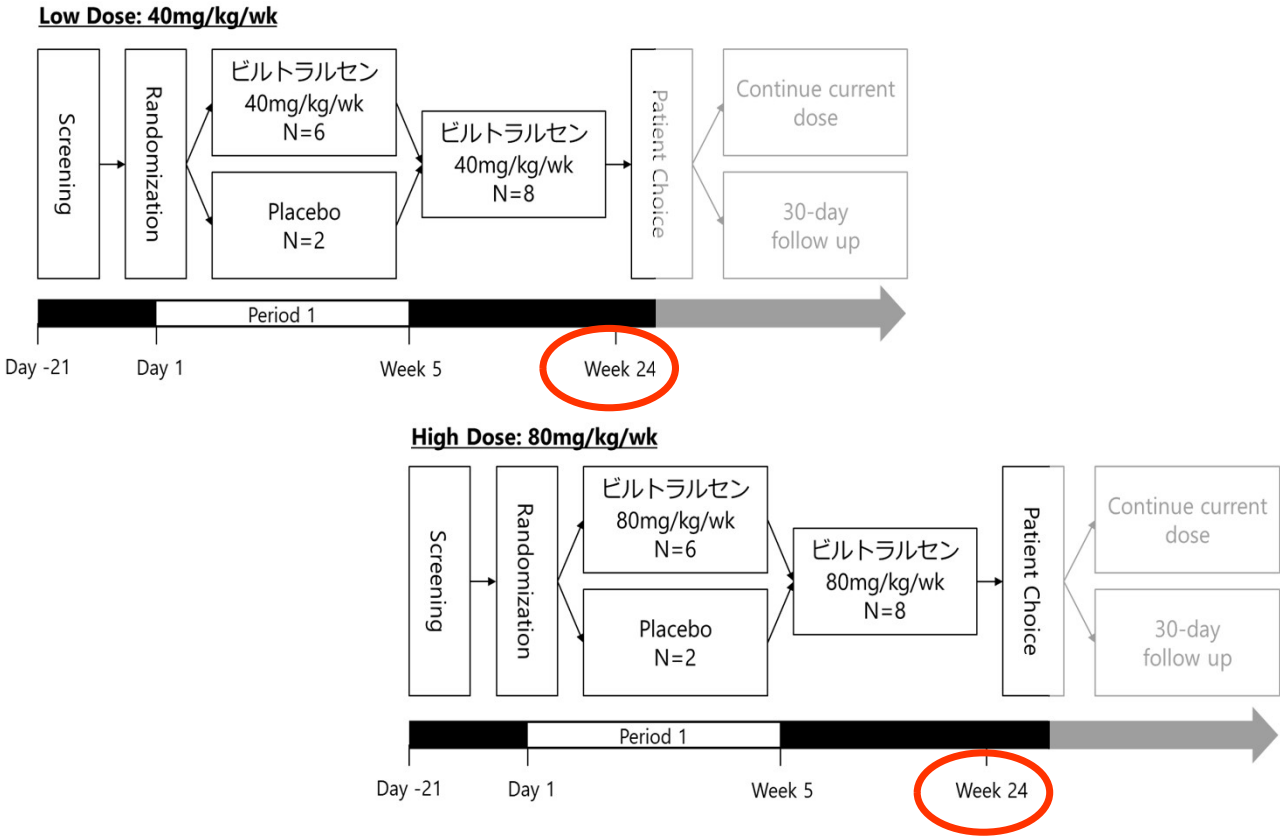


201試験　二重盲検試験

プラセボを対照においているが
投与後の有効性の評価は
Week 24で実施されるため
有効性の評価という観点では
事実上プラセボ対照ではない

主要評価項目はジストロフィン
タンパク発現

試験デザイン｜ビルトラルセンの試験

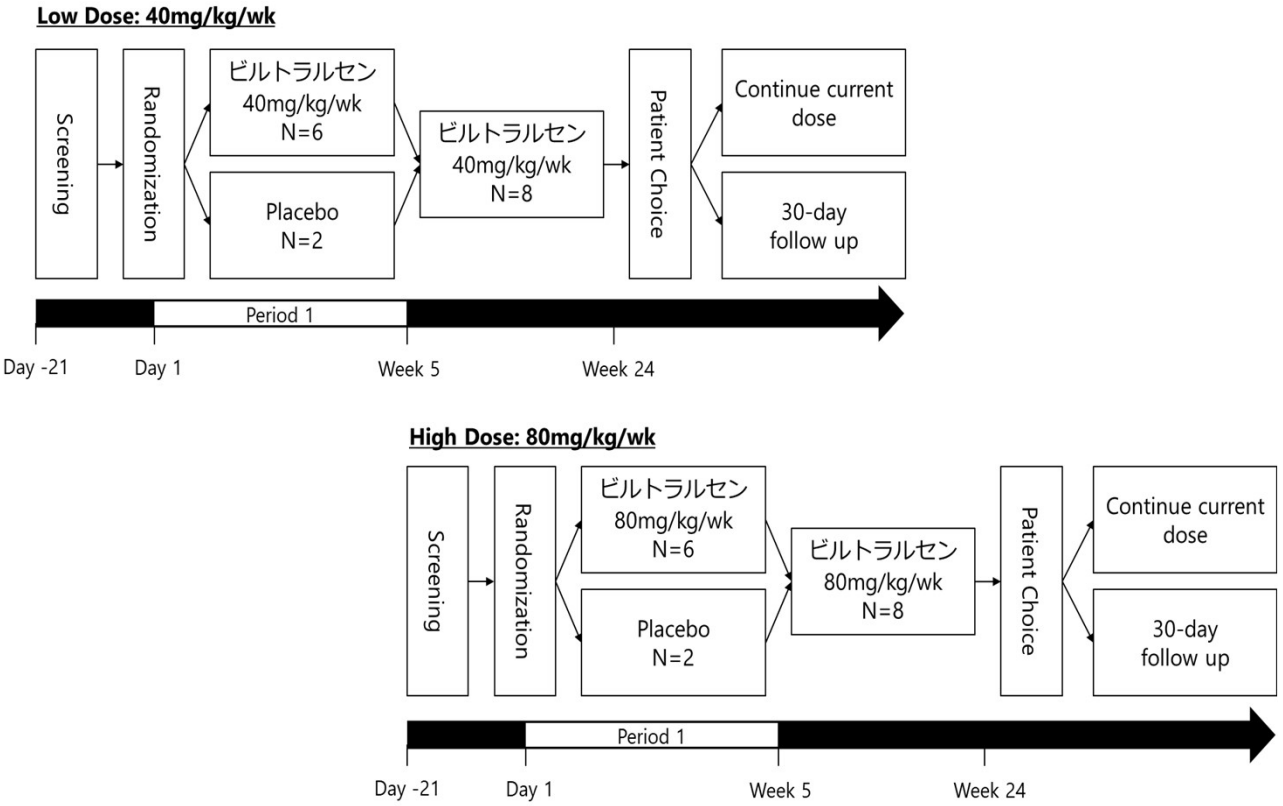


201試験　二重盲検試験

プラセボを対照においているが
投与後の有効性の評価は
Week 24で実施されるため
有効性の評価という観点では
事実上プラセボ対照ではない

主要評価項目はジストロフィン
タンパク発現

試験デザイン｜ビルトラルセンの試験



長期継続投与期

主要評価項目は 立ち上がり時間

運動機能評価の対照群に
CINRG[1]の自然歴集団を
ヒストリカルコントロールとして利用

[1] 米国の筋ジストロフィーの臨床試験ネットワークであるcooperative international neuromuscular research group （CINRG）

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[34]

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).

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[37]

	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)

PMDAからのコメント（ビルトラルセン）

審査報告書[34]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).

規制当局の反応が異なる要因として考えられるもの

製薬協

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[34]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).

まとめ｜ヒストリカルコントロールの利用

製薬協

試験実施前の検討の必要性・重要性

試験デザイン　Pocockの基準
実施手順、トレーニング方法、ヒストリカルコントロールと合わせる

解析計画　得られている情報や新規試験の目的に応じて解析手法を選択
解析手法の動作特性の把握
主解析の結果の安定性（ロバストネス）を評価する感度分析

事例から

ヒストリカルコントロールの利用では試験開始前に上記の検討が重要

Power prior、階層モデル、傾向スコアといった解析手法の選択を如何に事前に実施するかが課題

参考文献

[1] Ghadessi, Mercedeh, et al. "A roadmap to using historical controls in clinical trials-by Drug Information Association Adaptive Design Scientific Working Group (DIA-ADSWG)." Orphanet Journal of Rare Diseases 15.1 (2020): 1-19.

[2] 厚生労働省. "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." (2001).

[3] Food and Drug Administraion. "Rare Diseases: Natural History Studies for Drug Development (Draft)." (2019).

[4] Food and Drug Administraion. "Framework for FDA's Real-World Evidence Program." (2018).

[5] Chen, Jie, et al. "The current landscape in biostatistics of real-world data and evidence: clinical study design and analysis." Statistics in Biopharmaceutical Research (2021): 1-14.

[6] Pocock, Stuart J. "The combination of randomized and historical controls in clinical trials." Journal of chronic diseases 29.3 (1976): 175-188.

[7] 武田健太朗ら. "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 (2015): 25-50.

[8] Neuenschwander, Capkun-Niggli, et al. "Summarizing historical information on controls in clinical trials." Clinical Trials 7.1 (2010): 5-18.

[9] Gsteiger, Sandro, et al. "Using historical control information for the design and analysis of clinical trials with overdispersed count data." Statistics in Medicine 32.21 (2013): 3609-3622

参考文献

[10] Schmidli, Heinz, et al. "Robust meta-analytic-predictive priors in clinical trials with historical control information." Biometrics 70.4 (2014): 1023-1032.

[11] Hupf, Bradley, et al. "Bayesian semiparametric meta-analytic-predictive prior for historical control borrowing in clinical trials." Statistics in Medicine 40.14 (2021): 3385-3399.

[12] Hueber, Sands, et al. "Secukinumab, a human anti-IL-17A monoclonal antibody, for moderate to severe Crohn's disease: unexpected results of a randomised, double-blind placebo-controlled trial." Gut 61 (2012): 1693-1700.

[13] Food and Drug Administration. "Multi-disciplinary Review and Evaluation Benlysta® (belimumab) for Intravenous Infusion in Children 5 to 17 Years of Age with SLE. " (2019).

[14] Ibrahim, Chen. "Power prior distributions for regression models." Statistical Science (2000): 46-60.

[15] Rietbergen, Klugkist, et al. "Incorporation of historical data in the analysis of randomized therapeutic trials." Contemporary Clinical Trials 32. (2011): 848-855.

[16] Ollier, Morita, et al. "An adaptive power prior for sequential clinical trials - Application to bridging studies. " Stat Methods Med Res. 29.8. (2020): 2282-2294.

参考文献

[17] Duan, Yuyan, et al. "Evaluating water quality using power priors to incorporate historical information." Environmetrics: The Official Journal of the International Environmetrics Society 17.1 (2006): 95-106.

[18] Neuenschwander, Branson, et al. "A note on the power prior." Statistics in medicine 28.28 (2009): 3562-3566.

[19] Hobbs, Brian P., et al. "Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials." Biometrics 67.3 (2011): 1047-1056.

[20] Banbeta, Akalu, et al. "Modified power prior with multiple historical trials for binary endpoints." Statistics in Medicine 38.7 (2019): 1147-1169.

[21] Rosenbaum, Paul R., et al. "The central role of the propensity score in observational studies for causal effects." Biometrika 70.1 (1983): 41-55

[22] Liu, Meizi, et al. "Propensity-score-based meta-analytic predictive prior for incorporating real-world and historical data." Statistics in medicine 40.22 (2021): 4794-4808.

[23] Wang, Chenguang, et al. "Propensity score-integrated power prior approach for incorporating real-world evidence in single-arm clinical studies." Journal of Biopharmaceutical Statistics 29.5 (2019): 731-748.

参考文献

[26] BioMarin Pharmaceutical Japan 株式会社. “総合製品情報概要 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.” (2022).

[27] BioMarin Pharmaceutical Japan 株式会社. “セロイドリポフスチン症2型（CLN2）の治療と向き合うために.” (2021).

[28] Food and Drug Administration. “Brineura (cerliponase alpha) Injection Statistical Review and Evaluation.” (2017).

[29] Steinfeld, Robert, et al. "Late infantile neuronal ceroid lipofuscinosis: quantitative description of the clinical course in patients with CLN2 mutations." American journal of medical genetics 112.4 (2002): 347-354.

[30] BioMarin Pharmceutical Japan株式会社. “ブリニューラ脳室内注射液150mgに関する資料.”医薬品医療機器総合機構（2019）.

[31] Food and Drug Administration. “HIGHLIGHTS OF PRESCRIBING INFORMATION.” (2017).

[32] 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.

[33] ミクスOnline. ”米FDA　世界初のDMD治療薬Exondys51を承認.”

[34] 医薬品医療機器総合機構. “ビルテプソ点滴静注250 mg, 審査報告書.” (2020).

[35] Food and Drug Administration. “EXONDYS 51™ (eteplirsen) Statistical Review and Evaluation.” (2016).

参考文献

[36] 厚生労働省. “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.” (2001).
[37] 日本新薬株式会社. ビルテプソ点滴静注250 mgに関する資料 (2019)