

部会資料

# 医薬品開発における Model-based Meta Analysis の理論と実践

2026年05月  
Version 1.0

日本製薬工業協会  
医薬品評価委員会 データサイエンス部会

2025年度継続タスクフォース4

## 目次

目次	2
表一覧	3
図一覧	4
略語一覧	5
1 はじめに	8
2 MBMA の方法論	10
2.1 理論背景	10
2.1.1 基本となる考え方	10
2.1.2 変量効果の階層	11
2.1.3 薬効・用量の組み込み	12
2.1.4 時系列 (Longitudinal) 解析	13
2.2 モデル構築のためのデータ収集・データベースの構築	15
2.2.1 データベース構築のプロセスとハードル	15
2.2.2 データのデジタルイズ	17
2.2.3 AI 活用によるデータ収集	17
2.2.4 MBMA に必要な試験数	18
2.2.5 Publication バイアスの評価	18
2.2.6 異なる補完方法を用いたデータへの対応	19
2.3 MBMA モデルの構築	21
2.3.1 Landmark 解析と時系列解析のすみ分け	21
2.3.2 変量効果構造の設定 (試験間差, 群間差, 個体差, 残差)	22
2.3.3 時点間の相関の取り扱い	22
2.3.4 残差の重みづけ	23
2.3.5 共変量探索	24
2.3.6 異質性の評価と対応	24
2.3.7 パラメータ推定方法とソフトウェア	25
2.3.8 モデル評価・妥当性確認	26
2.3.9 要約データと個別データの統合	27
2.3.10 臨床試験のエンドポイントが連続変数でない場合	28
2.4 研究の質の担保	28
3 MBMA モデルの活用	30
3.1 モデル活用の概要	30
3.2 個別事例-1 開発早期: 適切なバイオマーカーの選定	33
3.3 個別事例-2 検証試験前: 用量設定試験の代替	37
3.4 個別事例-3 検証試験前: Go/no-go 判断のための競合薬との比較	42
3.5 個別事例-4 併用戦略: 安全性プロファイルの評価	45
4 まとめ	49
5 引用文献	51

## 表一覧

表 1	開発相ごとのリサーチクエスチョンに対する MBMA の活用例.....	31
表 2	個別事例-1（開発早期：適切なバイオマーカーの選定）の概要.....	33
表 3	個別事例-2（検証試験前：用量設定試験の代替）の概要.....	37
表 4	個別事例-3（検証試験前：Go/no-go 判断のための競合薬との比較）の概要.....	42
表 5	個別事例-4（併用戦略：安全性プロファイルの評価）の概要.....	45

## 図一覧

図 1	医薬品開発における MBMA の適用範囲.....	9
図 2	固定効果の仮定の図示 .....	10
図 3	変量効果の仮定の図示 .....	11
図 4	固定効果/変量効果の階層の図示.....	12
図 5	メタアナリシスのためのデータベース構築のフローチャート .....	15
図 6	PRISMA2020 のフローダイアグラムのテンプレート.....	16
図 7	ファンネルプロットの例.....	19
図 8	メタアナリシスのためのモデル構築のフローチャート .....	21
図 9	臨床アウトカム（重症度レベル）の分類.....	38
図 10	PD-1 阻害薬（左図）又は CTLA-4 阻害薬（右図）の曝露量と Grade 3/4 の治療関連有害事象（trAE）の発現率の関係 .....	47
図 11	MBMA のプロセスの概観と各プロセスに関連する主なトピック .....	50

## 略語一覧

略語	正式名称	日本語
ACR20/50/70	American College of Rheumatology 20/50/70 response criteria	ACR20/50/70 反応率
ADAS-cog	Alzheimer's Disease Assessment Scale-cognitive subscale	アルツハイマー病評価尺度
AE	Adverse Event	有害事象
AI	Artificial Intelligence	人工知能
AIC	Akaike Information Criterion	赤池情報量基準
AMSTAR-2	A Measurement Tool to Assess Systematic Reviews 2	—
AR(1)	First-order Autoregressive	一次自己回帰構造
AUC	Area Under the Curve	濃度-時間曲線下面積
BID	Bis In Die	1日2回投与
BOCF	Baseline Observation Carried Forward	ベースライン値補完法
BTK	Bruton's Tyrosine Kinase	ブルトン型チロシンキナーゼ
CFB	Change From Baseline	ベースラインからの変化量
CFFT	Critical Flicker Fusion Test	臨界融合頻度試験
CS	Compound Symmetry	複合対称構造
CTCAE	Common Terminology Criteria for Adverse Events	有害事象共通用語規準
CTLA-4	Cytotoxic T-Lymphocyte Antigen 4	細胞傷害性 T リンパ球抗原 4
DSST	Digit Symbol Substitution Task	数字記号置換課題
E-R	Exposure-Response	曝露-反応関係
EC50	50% of Effective Concentration	50%有効濃度
ED50	50% of Effective Dose	50%有効用量
EM	Expectation-Maximization	期待値最大化法
Emax	Maximum Effect	最大効果
FOI	Force of Infection	感染力
GABA	Gamma-Aminobutyric Acid	γ-アミノ酪酸
GABAA	Gamma-Aminobutyric Acid Type A receptor	GABA A 受容体
ICH	International Council for Harmonisation	医薬品規制調和国際会議
ICI	Immune Checkpoint Inhibitor	免疫チェックポイント阻害薬
ICU	Intensive Care Unit	集中治療室
imAE	Immune-related Adverse Event	免疫関連有害事象
IPD	Individual Participant/Patient Data	個別患者データ
IT50	50% Inhibitory Titer	50%阻害力価
LOAO	Leave-One-Arm-Out	—

略語	正式名称	日本語
LOCF	Last Observation Carried Forward	最終観測値引き継ぎ法
LOSO	Leave-One-Study-Out	—
LRTI	Lower Respiratory Tract Infection	下気道感染
M&S	Modeling & Simulation	—
mAb	Monoclonal Antibody	モノクローナル抗体
MBMA	Model-Based Meta-Analysis	モデルに基づいたメタアナリシス
MCMC	Markov Chain Monte Carlo	マルコフ連鎖モンテカルロ法
MeSH	Medical Subject Headings	—
MIDD	Model-Informed Drug Development	モデルに基づく医薬品開発
MIPD	Model-Informed Precision Dosing	モデルに基づく個別化投与
MMRM	Mixed Model for Repeated Measures	反復測定混合効果モデル
MTX	Methotrexate	メトトレキサート
MTX-IR	Methotrexate Inadequate Response	メトトレキサート効果不十分
NMA	Network Meta-Analysis	ネットワークメタアナリシス
NONMEM	Nonlinear Mixed Effects Modeling	非線形混合効果モデリング
NSCLC	Non-Small Cell Lung Cancer	非小細胞肺癌
PASI	Psoriasis Area and Severity Index	乾癬面積重症度指数
PBPK	Physiologically Based Pharmacokinetic	生理学的薬物速度論
PD	Pharmacodynamics	薬力学
PD-1	Programmed Death-1	プログラム細胞死 1
PD-L1	Programmed Death-Ligand 1	プログラム細胞死リガンド 1
PICO	Population, Intervention, Comparison, Outcome	対象・介入・比較・評価項目
PK	Pharmacokinetics	薬物動態
PK/PD	Pharmacokinetics/Pharmacodynamics	薬物動態/薬力学
PMx	Pharmacometrics	ファーマコメトリクス
PoC	Proof of Concept	概念実証
popPK	Population Pharmacokinetics	母集団薬物動態
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses	—
Q-Q	Quantile-Quantile	—
QD	Quaque Die	1日1回投与
RA	Rheumatoid Arthritis	関節リウマチ
RCT	Randomized Controlled Trial	ランダム化比較試験
RMSE	Root Mean Square Error	二乗平均平方根誤差
RSV	Respiratory Syncytial Virus	RS ウイルス

略語	正式名称	日本語
SD	Standard Deviation	標準偏差
SE	Standard Error	標準誤差
SEM	Saccadic Eye Movement	衝動性眼球運動
SNA	Serum Neutralizing Antibody	血清中和抗体
STAR-MBMA	Standards for Transparent and Accurate Reporting of Model-Based Meta-Analyses	—
TNF-IR	Tumor Necrosis Factor Inadequate Response	TNF 阻害薬効果不十分
trAE	Treatment-related Adverse Event	治療関連有害事象
UN	Unstructured	無構造共分散
VAS	Visual Analogue Scale	視覚的アナログ尺度
VPC	Visual Predictive Check	視覚的予測チェック
WOCF	Worst Observation Carried Forward	最悪値補完法

—：適切な日本語訳なし

## 1 はじめに

医薬品開発は不確実性が高く、成功確率の低下と開発コストの高騰が大きな課題となっている[1, 2]。医薬品開発は連続的かつ反復的な学習プロセスであり、情報の限られた試験結果を将来の意思決定への確に反映するために、外部エビデンスを含む総合的な根拠にもとづく意思決定が不可欠である。この課題に対する一つのアプローチとして、数理モデルを用いたシミュレーション (Modeling & Simulation : M&S) 及び、それに基づいた開発戦略 Model-Informed Drug Development (MIDD) が提唱されている[3]。MIDD で扱われる一つのアプローチとして、モデルを用いたメタアナリシス Model-Based Meta Analysis (MBMA) がある。MBMA は文字通り、数理モデルに基づいてメタアナリシスを行う手法であり、従来のメタアナリシスが単一用量・単一時点の平均効果を中心に複数のソースを統合解析するのに対し、MBMA では必要に応じて用量反応、曝露反応、時間経過、そしてエンドポイント間の関連をモデル化し、目的に応じて解析の対象を拡張可能な点に特徴がある[4, 5]。医薬品開発においては、自社開発品由来のデータのみならず、比較対象となるプラセボ・標準治療の効果、同じ薬効メカニズムを持つ競合薬の反応性や時間推移、地域や背景治療の違いによる反応への影響など、外部エビデンスから取得可能な情報も意思決定を行う上で、重要な根拠となる[6, 7]。とりわけ、このような情報に基づいた MBMA の結果は、導入品等の Due Diligence においても重要なベンチマークとしての役割を果たしている。こうした状況で MBMA は、統計学的な厳密性を保ちつつ薬理的な妥当性を組み込み、社内外の要約データと(入手可能な場合)個別データを統合して、意思決定に直結するシミュレーションを可能にする枠組みであり、医薬品の曝露-反応解析ガイドライン[8]や ICH M15 (General Principles for Model-Informed Drug Development M15) [9]でも取り上げられ、近年報告数が顕著に伸びているアプローチである[10, 11]。

MBMA は、探索研究から臨床開発、さらに市販後の最適化に至るまで、医薬品開発の広範な段階に適用可能である(図1)。例えば、開発の初期段階では、作用機序が同じ他剤を対象に、用量反応や曝露反応の MBMA モデルを構築し、ヒト有効曝露の目標値を推定して化合物候補の選択や初期用量の設定を支援できる[12, 13]。臨床第2相試験から第3相試験への移行では、類効薬や標準治療のデータを統合した MBMA モデルを事前に構築しておくことで、自社開発品の第2相試験の結果が得られた直後に第3相試験結果を多様な比較薬や対象集団の条件でシミュレーションし、自社開発品の類効薬に対するベンチマーキング、成功確率、必要症例数、評価時点の妥当性を定量的に評価できる[14, 15]。また類効薬のデータに基づいて、バイオマーカーなどの短期エンドポイントから臨床評価項目などの長期アウトカムを繋げるモデルを構築すれば、自社開発品の早期探索試験での短期エンドポイントの結果に基づいて、後期試験の長期アウトカムを予測し、意思決定を加速することも可能になる[16, 17]。さらに後期開発では、Head to Head の比較試験がない比較対象との比較においてネットワークメタアナリシスを活用することがあるが[18]、ここに MBMA で用量反応や時間推移を組み込むことにより、治療間比較の拡張性と精度を高めることが可能となる[19]。

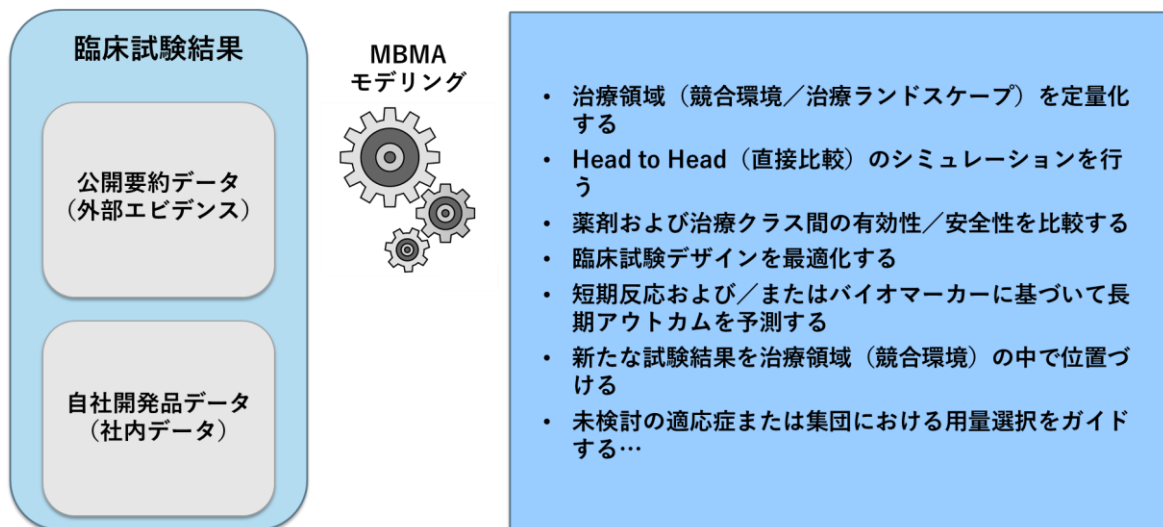


図 1 医薬品開発における MBMA の適用範囲

Simeoni M et al (2022)[20]より作成

MBMA を実施する上では、MIDD で扱われる他のアプローチとは異なる幾つの特徴がある。まず、メタアナリシス全般に言える特徴として、解析対象となるデータベース構築の難易度が高い点である。上述の通り、MBMA の対象となるデータは公表論文、学会発表、規制当局レビュー、試験登録情報といった外部エビデンスが主となり、これらはデータソースによって情報の粒度が異なり、また整合性が取れていない場合が多い（例えば、ある報告ではエンドポイントの平均値と標準偏差が数値として表で報告され、別の報告では平均値±標準誤差のプロファイルの図のみが報告されている）。これらを統合して解析するにあたっては、データの取得方法・選択基準を事前に定義し、情報が欠測した場合の対応について検討する必要がある。また、ポジティブな結果が公表されやすい（出版バイアス）という公表資料特有の構造的リスクも存在する。さらに、異なる情報（要約データ [外部エビデンス] と個別データ [自社開発品のデータ]）を同時に扱うケースがあり、その場合の取り扱い方・解釈にも注意が必要な点が多くある。これらのメタアナリシス全般に言える特徴に加えて MBMA では、用量反応の組み込み方、時点間相関の取り扱い、共変量効果の取り扱い、変量効果構造の設定、残差の重みづけなどのモデル構造の設定にも検討を要する。このように、MBMA では自社開発品の社内データのみを扱う解析とは異なる pit fall が多数存在する。一方で、これまでのところ MBMA について日本語で詳しく解説された書籍・総説は存在しない。

以上を踏まえ、本報告書は、医薬品開発における MBMA の理論と応用について、主に既報告の例を参照して整理し、社内の意思決定において、MBMA を活用するための導入的な資料となること目指している。本報告書は主に二つのセクションに分かれており、前半のセクション (2 項) は主に MBMA の実施担当者となる統計解析部門や臨床薬理部門の担当者を想定読者とし、MBMA の理論について、その方法論と特徴的なトピックをまとめている。後半のセクション (3 項) は MBMA を戦略的に用いる開発担当者全般を想定読者とし、MBMA の医薬品開発での実践を実例を交えて紹介する。

## 2 MBMA の方法論

本セクションでは MBMA の方法論について、その理論的な背景と特徴的なトピックをまとめて解説する。

### 2.1 理論背景

#### 2.1.1 基本となる考え方

MBMA では複数の試験データを統合解析して、興味の対象 (=真の薬剤/治療効果量) を二つの効果の組み合わせ=固定効果 (Fixed effect) + 変量効果 (Random effect) でモデル化する。

固定効果は一般にモデル内でその効果が一定 (固定) 値と見なされる要素で、MBMA においては、試験間や個体間等での変動の影響を受けず、常に一定となる値を表すパラメータである。ここで、複数の試験の効果量を統合して解析する場合、解析対象データに偏りがなければ [図 2](#) に示すように各試験で観察される治療効果 ( $Y_i$ ) は真の効果量  $\theta$  の周辺に均等にばらつくと想定される。この真の効果量  $\theta$  は MBMA で興味の対象となる代表的な固定効果である。

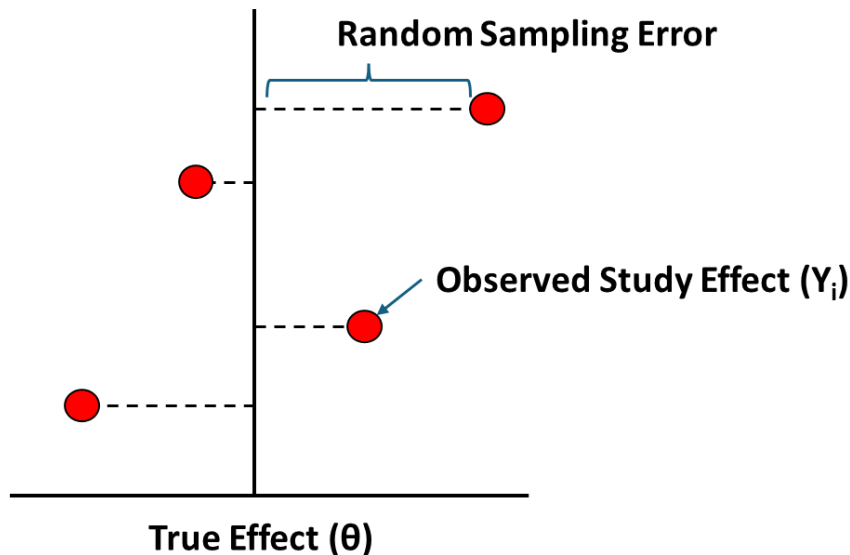


図 2 固定効果の仮定の図示

Boucher M et al (2016)[\[4\]](#)より作成

一方、実際に観察される  $Y_i$  は試験ごとの患者集団やデザインの違いに由来して  $\theta$  からは乖離した値 ([図 3](#),  $\theta_i$ ) をとり、さらに仮に同じ患者集団・デザインで試験を実施した場合でも、観測値はランダムなサンプリング誤差によりばらつく ([図 3](#) 右)。このように、モデル内のばらつきを表す値を変量効果と呼ぶ。MBMA においては試験間差や群間差、その他のばらつき (残差変動) がその大きさを表すパラメータである。なお後述 ([2.3 項](#)) の通り、モデル構築にあたっては、まずデータの特徴に基づいて変量効果のモデル構造を検討し、その後必要な固定効果の組み込みを検討する手順が推奨されている。

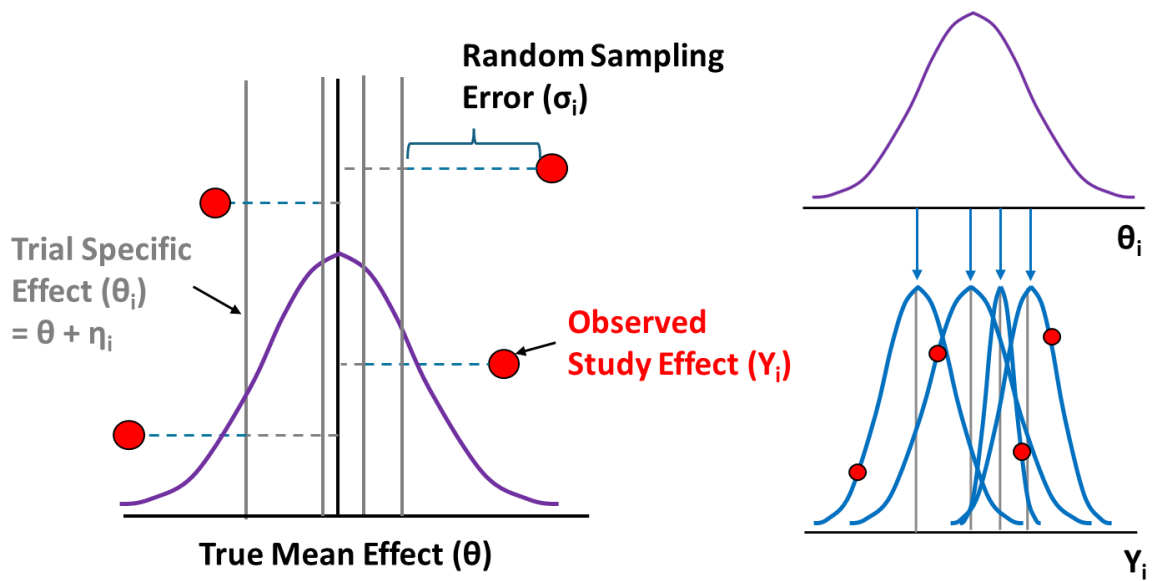


図 3 変量効果の仮定の図示

Boucher M et al (2016)[4]より作成

これら固定効果，変量効果の関係を数式に書き下すと，以下のように記載される。

$$Y_i = \theta + \eta_i + \sigma_i, \quad \eta_i \sim N(0, \Omega), \quad \sigma_i \sim N(0, \Sigma)$$

ここで， $Y_i$ は試験*i*の治療効果， $\theta$ は真の治療効果（固定効果）， $\eta_i$ は試験間差（変量効果）， $\sigma_i$ は残差変動（変量効果）となる。なお解析の対象となるエンドポイントがイベントの発生の有無などの2値データや，time to event型のデータである場合，これらのパラメータをロジットスケールやハザード関数として定義して解析する。

### 2.1.2 変量効果の階層

MBMAでは変量効果をばらつきの種類に基づいて，さらに細分化・多層化する。[図4](#)にMBMAで取り扱われる代表的な変量効果階層をまとめた。

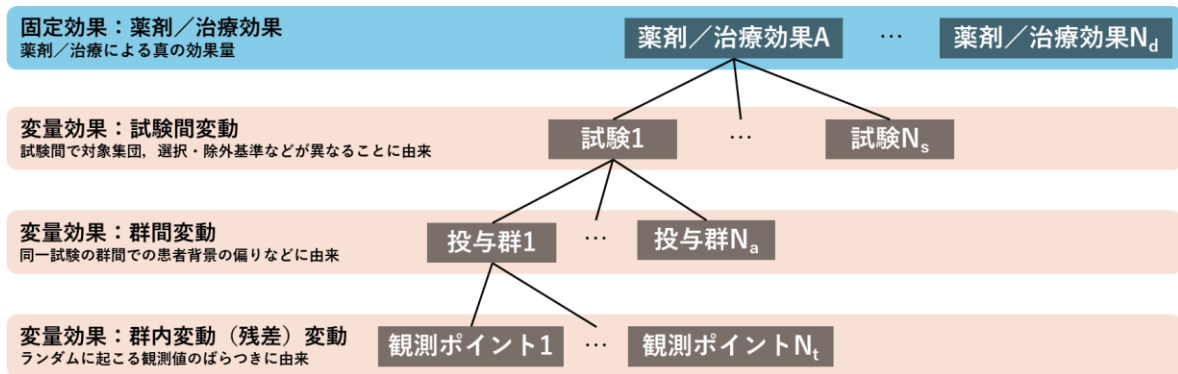


図 4 固定効果/変量効果の階層の図示

荒野俊平ら (2015) [21]より作成

これらの異なる変量効果を数式化した MBMA モデルは以下となる。

$$\begin{aligned}
 Y_{ikt} &= \theta + \eta_i^{study} + \eta_{ik}^{arm} + \delta_{ikt} \\
 \eta_i^{study} &\sim N(0, \Omega^{study}) \\
 \eta_{ik}^{arm} &\sim N(0, \Omega^{arm}) \\
 \delta_{ikt} &\sim N(0, \Sigma)
 \end{aligned}$$

ここで、 $Y_{ikt}$  は試験  $i$ 、群  $k$ 、時点  $t$  での治療効果、 $\eta_i^{study}$  は試験  $i$  の試験間差を表す変量効果、 $\eta_{ik}^{arm}$  は試験  $i$ 、群  $k$  の群間差を表す変量効果、 $\delta_{ikt}$  は群内（もしくは残差）変動であり、 $\Omega$  及び  $\Sigma$  は変量効果の分布の分散を表す。なお、収集したデータによってはこれらの 3 階層を全て推定することが出来る訳ではない点に留意すべきである。例えば、収集した全ての試験が単群の試験であった場合、群間変動と群内変動は分離評価できないため、試験間変動と残差変動の 2 階層を設定することとなる（詳しくは [2.3.2 項](#) を参照）。

また、重要な点として、これらの変量効果のうち、群間変動及び群内変動は観測された標準誤差や例数によって影響を受けるため（1 群当たりの例数が多ければ値の偶発的な偏りは生じにくくなり、精度が高い）、それらの大小に基づいた補正が行われる。この補正にはいくつかの方法が存在するが（[2.3.4 項](#) 参照）、シンプルには例数の平方根で補正するアプローチがある。これを適用した場合、上記の式は以下となる。

$$Y_{ikt} = \theta + \eta_i^{study} + \frac{1}{\sqrt{N_{ik}}} * \eta_{ik}^{arm} + \frac{1}{\sqrt{N_{ik}}} * \delta_{ikt}$$

$N_{ik}$  : 試験  $i$ 、群  $k$  での例数

### 2.1.3 薬効・用量の組み込み

薬剤の影響を検討する上では、下記のようにプラセボを含む薬剤の影響を、固定効果として別途推定する。試験間でプラセボの結果にばらつきがあるような場合は、変量効果又は共変量では説明しきれないことがあるため、試験ごとのプラセボの効果を局外パラメータとして設定し、プラセボ補正後の結果に着目して解析することがある[19]。また複数の薬剤が解析対象に含まれる

場合、薬剤の影響のパラメータは薬剤分設定することとなる。

$$Y_{ik} = \theta_{E0} + \theta_{Drug1} + \theta_{Drug2} \dots + \eta_i^{study} + \frac{1}{\sqrt{N_{ik}}} * \eta_{ik}^{arm} + \frac{1}{\sqrt{N_{ik}}} * \delta_{ikt}$$

$\theta_{E0}$  : プラセボ効果,  $\theta_{DrugX}$  : Drug X の薬効

さらに同一薬剤でも異なる用量のデータが解析に含まれる場合、その影響も薬効モデルの固定効果として組み込まれる。用量反応の特徴に基づいて、薬効モデルの記述式は選択すべきだが、代表的な Emax モデルを採用した場合、以下の式となる。変量効果は ED50 等にも組み込むことが理論上可能であるが、この場合、ED50 の試験間差等を推定する上で十分な用量幅やデータ数が得られているのか、吟味する必要がある。

$$Y_{ik} = \theta_{E0} + \frac{Emax_{Drug1} * Dose_{Drug1}}{ED50_{Drug1} + Dose_{Drug1}} + \frac{Emax_{Drug2} * Dose_{Drug2}}{ED50_{Drug2} + Dose_{Drug2}} + \dots$$

$$+ \eta_i^{study} + \frac{1}{\sqrt{N_{ik}}} * \eta_{ik}^{arm} + \frac{1}{\sqrt{N_{ik}}} * \delta_{ikt}$$

$Emax_{DrugX}$  : Drug X の最大効果,  $ED50_{DrugX}$  : Drug X の最大効果の 50% を発揮する用量,  
 $Dose_{DrugX}$  : Drug X の用量

なお同一作用機序の薬剤や、得られたデータから最大薬効が同じと想定できる場合、上記の Emax は薬剤間で統一することも考えられる。Emax モデルの他にも、データの特性に依じて対数線形 (log-linear)、二次 (quadratic)、指数関数 (exponential) などが用いられる。

### 2.1.4 時系列 (Longitudinal) 解析

薬効の発現時間 (オンセット) や持続時間など、時系列な薬効の推移に興味の対象がある場合、対象となる全時点での測定点を解析対象とし、時点の影響をモデルの中に固定効果として組み込む。薬効の時間推移が何によるものなのか (疾患の自然悪化、プラセボ効果、血中濃度の蓄積、ターゲットのターンオーバーなど) を仮定し、時点の影響を組み込むパラメータを検討する。例えば、プラセボ効果にのみ時間の影響 (時間に伴って徐々にプラセボ効果が発現) が認められる場合には例として、以下の様なモデルが挙げられる。

$$Y_{ik} = \theta_{E0} * (1 - e^{-k*time}) + \frac{Emax_{Drug1} * Dose_{Drug1}}{ED50_{Drug1} + Dose_{Drug1}} + \frac{Emax_{Drug2} * Dose_{Drug2}}{ED50_{Drug2} + Dose_{Drug2}} + \dots$$

$$+ \eta_i^{study} + \frac{1}{\sqrt{N_{ik}}} * \eta_{ik}^{arm} + \frac{1}{\sqrt{N_{ik}}} * \delta_{ikt}$$

$k$  : プラセボ効果の発現速度定数

一方で、薬効発現に時間の影響遅れがある場合は、以下のように表現できる。

$$\begin{aligned}
Y_{ik} &= \theta_{E0} + \frac{Emax, T_{Drug1} * Dose_{Drug1}}{ED50_{Drug1} + Dose_{Drug1}} + \frac{Emax, T_{Drug2} * Dose_{Drug2}}{ED50_{Drug2} + Dose_{Drug2}} + \dots \\
&\quad + \eta_i^{study} + \frac{1}{\sqrt{N_{ik}}} * \eta_{ik}^{arm} + \frac{1}{\sqrt{N_{ik}}} * \delta_{ikt} \\
Emax, T_{Drug1} &= \frac{Emax_{Drug1} * time}{ET50_{Drug1} + time} \\
Emax, T_{Drug2} &= \frac{Emax_{Drug2} * time}{ET50_{Drug2} + time}
\end{aligned}$$

$Emax, T_{DrugX}$  : time=T での Drug X の最大効果,  $ET50_{DrugX}$  : Drug X の最大効果の 50%を發揮する時間

時系列解析は一時点を使った解析 (Landmark 解析) に比べて多くの情報を得ることが出来るが, モデルの複雑さが増し, 仮定も多くなるため, 後述の通りその使い分けについては留意すべきである。

## 2.2 モデル構築のためのデータ収集・データベースの構築

### 2.2.1 データベース構築のプロセスとハードル

メタアナリシスは、複数の研究結果を統合し、より信頼性の高い結論を導くための手法であり、その成功には質の高いデータベースの構築が不可欠である。データベース構築のため、体系的かつ再現性のあるプロセスが必要となる。[図 5](#) にデータベース構築に必要なフローを示し、各プロセスとハードルを説明する。

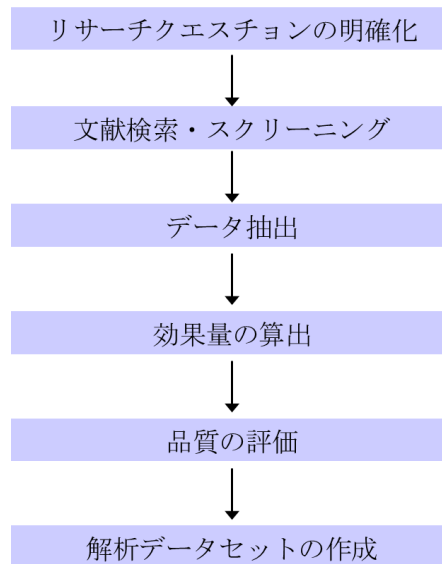


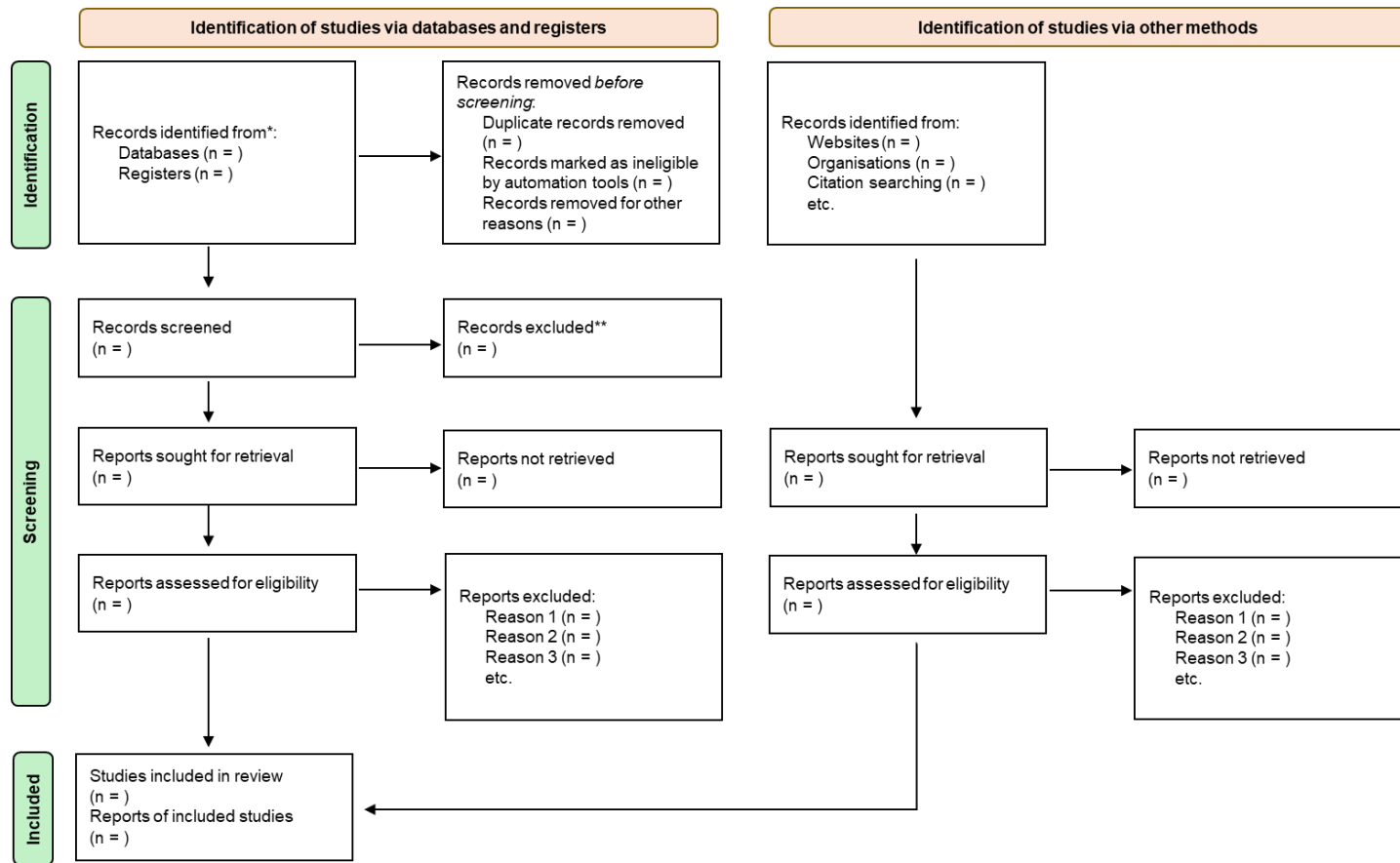
図 5 メタアナリシスのためのデータベース構築のフローチャート

#### ① リサーチクエスチョンの明確化

最初に研究の主題を明確化し、PICO (Population [対象], Intervention [治療などの介入], Comparison [比較対照], Outcome [評価指標]) を定義して、体系的レビューの範囲を決める。

#### ② 文献検索・スクリーニング

関連する研究を PubMed や Embase などの複数のデータベースを活用し、包括的な文献検索を行う。検索式にはキーワードと MeSH (米国国立医学図書館が作成した統制語彙) を組み合わせ、学会抄録や未出版データも可能な限り収集する。取得した文献は重複を除去し、事前に設定した選定基準に基づいて、タイトル→アブストラクト→フルテキストの順でスクリーニングを実施する。フルテキストのスクリーニングでは少なくとも 2 人のレビュアーによる独立評価を行い、バイアスを低減することが推奨されている。システムティックレビューやメタアナリシスにおいて、文献がどのような過程で特定・選別・除外・最終採択されたかを文献数とともに可視化するため、PRISMA 2020 ([2.4 項](#)参照) でフローダイアグラムのテンプレートが提示されている ([図 6](#))。



\*Consider, if feasible to do so, reporting the number of records identified from each database or register searched (rather than the total number across all databases/registers).  
 \*\*If automation tools were used, indicate how many records were excluded by a human and how many were excluded by automation tools.

図 6 PRISMA2020 のフローダイアグラムのテンプレート

### ③ データ抽出

選定基準 (対象集団, 薬剤や治療の種類, 比較対照など) を満たした研究から, 標準化されたフォームを用いてデータ抽出を行い, 著者情報, 症例数, 介入内容, アウトカム, 統計値などを整理する。主観的な解釈が必要となる情報や, 結果の解釈に重要となる情報 (例: アウトカムデータ) は, 少なくとも2名の担当者が独立して抽出することが推奨されている。

### ④ 効果量の算出

平均差, 標準化平均差, オッズ比, リスク比など, 解析に適した効果量へ変換する。

### ⑤ 品質の評価 (Risk of Bias 評価)

個々の研究の質が結果に影響するため, 各研究がどれだけバイアスの影響を受けているか体系的に評価する。Cochrane ハンドブックで推奨されている Risk of Bias tool などを用いて評価を行う。

### ⑥ 解析用データセットの作成

解析ソフトウェアに取り込める形式に整理し, 解析用データセットを作成する。

以上のデータベース構築のプロセスにはいくつかのハードルがある。検索戦略の不備による重要な研究データの漏れ, 複数論文にまたがる重複データの処理不足, 報告の不完全性による統計値不足, 介入やアウトカムの異質性, 品質評価の主観性などが挙げられる。これらは解析全体のバイアスにつながる可能性がある。

Cochrane ハンドブック (Cochrane Handbook for Systematic Reviews of Interventions)[23]は, 体系的レビューの実施に関する包括的なガイドラインであり, 研究の計画, 文献検索, データ抽出と評価, 解析, 結果の解釈と報告における指針を提供している。このガイドラインは, 質の高い体系的レビューとメタアナリシスを行うための標準として広く認識されており, 研究者にとって信頼性の高い情報源となっている。

## 2.2.2 データのデジタイズ

本来論文に掲載されている数値データ (平均値・標準偏差など) を直接取得することが望ましいが, 実際には論文中にデータがグラフでのみ表示されている場合がある。このようなときに, 図から目盛りを読み取り, 元の数値を推定することができる。このような数値の推定を含め, アナログ情報や物理的なデータをデジタル形式に変換するプロセスをデジタイズ (digitize) という。デジタイズのツールとして, WebPlotDigitizer, Engauge Digitizer, ycasd, Plot Digitizer, GetData Graph Digitizer などが知られている (WebPlotDigitizer, Engauge Digitizer, Dexter 及び ycasd はオープンソースとして, Plot Digitizer [一部の機能は無料] 及び GetData Graph Digitizer は商用ソフトウェアとして提供されている)。ソフトウェアは視覚的な推定や定規の使用よりも便利で正確であることが示されている。

## 2.2.3 AI 活用によるデータ収集

人工知能 (AI) 技術の高度化に伴い, 従来は人手に依存していた文献検索, データ抽出 (上記のデジタイズによるデータ抽出を含む), PDF 解析, グラフ読取, Web 情報収集などの作業に AI を

導入することで、作業効率の向上、ヒューマンエラーの低減、解析の迅速化、及び大量データを扱う研究や実務において効率的な情報収集が可能となることが期待される。AIを活用したデータ収集フローとして以下が挙げられる。

- AIによる論文検索の最適化（Elicit等）
- AIによる文献スクリーニング補助（タイトル・アブストラクトの分類）
- PDF / 画像からの自動データ抽出
- 図表のAIデジタイズによる数値化
- 抽出データの自動構造化（Excel / CSVへ変換）
- 統計ソフト（R・RevMan等）への取り込み

AI活用における注意点として、AIが抽出したデータには、数値の転記ミスや抽出漏れといった誤りに加え、内容の誤解や不正確な解釈が含まれる可能性がある。そのため、専門家による確認（Human-in-the-loop）が不可欠である。医療データや企業内部情報を扱う場合は、AIプラットフォームのセキュリティ要件や遵守すべきガイドラインの確認が必要である。また、個人情報保護・著作権への配慮も必要である。

#### 2.2.4 MBMAに必要な試験数

MBMAに必要な試験数を明確に規定したガイドラインはなく、公表論文も限られている。またMBMAの目的によって、必要な試験数は異なることが予想される。McLellanらはCochraneデータベースを用いた検討において9試験あれば、フルのメタアナリシスとの比較において80%が一致することや推定された効果の相関係数が0.96であることを報告している[24]。また、Boucherらは試験数が少ないとき（例えば5試験未満）、試験間誤差の推定精度が十分ではないことを述べている[4]。

#### 2.2.5 Publicationバイアスの評価

メタアナリシスにおける主要な懸念の一つとして、すべての研究が公表されているとは限らない点が挙げられる。例えば、主要目的が達成できなかった試験や途中で中止された試験では、試験結果が公表されていない場合が想定される。このため、公表された結果のみに基づく統合解析では、ポジティブな方向へのバイアスが生じる可能性がある。したがって、publicationバイアスの評価はメタアナリシスにおいて重要である。Publicationバイアスの視覚化ツールとして、funnelプロットが有用である（図7）。Funnelプロットは、x軸に治療効果の推定値（群間差、リスク比等）、y軸に推定精度の指標（サンプルサイズや推定値の標準誤差等）をプロットしたものである。全く同じ条件で繰り返し評価された結果は真値を中心に対称にばらつき、推定精度が高い研究結果は真値に近い値となり、推定精度が低い研究では真値を中心に広く結果がばらつく。したがって、publicationバイアスが小さく研究の数が多ければ、プロットのばらつき具合が漏斗（funnel）を逆さにした左右対称型の形状を示す。ただし、解析に含まれる研究数が少ない場合には、対称性・非対称性を明確に判断することは困難であり、またプロットは偶然のばらつきや、研究デザインや対象集団など研究間の差異（異質性）といった複数の要因の影響を受けることには留意が必要である。

また、公表された試験結果であっても、文献では主要評価時点のみの結果が報告され、その他の時点の結果が報告されていない場合も考えられる。MBMAにより経時的な解析をする場合は、この点も留意が必要であり、各時点においてどの試験が寄与しているのかを理解することも重要である。また、時点ごとに funnel プロットを作成し評価するアプローチも有用かもしれない。

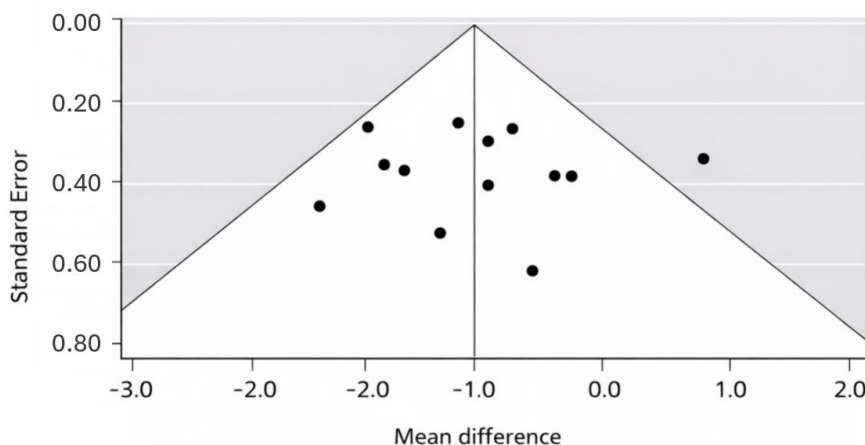


図 7 ファンネルプロットの例

メタアナリシスの解析対象とする研究 (study) の治療効果の推定値 (群間差, リスク比等) を横軸, 推定精度 (サンプルサイズや推定値の標準誤差等) の指標を縦軸にプロットする。Publication バイアスが小さく研究の数が多ければ, プロットは左右対称の漏斗状の形状を示す。

### 2.2.6 異なる補完方法を用いたデータへの対応

MBMAに限らずメタアナリシスにおける課題の一つとして、統合対象の臨床試験間で異なる方法で解析されていることが挙げられる。ICH E9 (R1) で示されている estimand の枠組みに照らすと、各試験ではそれぞれ固有の estimand (推定する治療効果) が設定され、解析対象の試験間で estimand の構成要素 (患者集団、治療、中間事象とそれらの解析上での取り扱い等) が異なる状況も想定される。試験間で estimand を完全に揃えることは、特に個別患者データへのアクセスが限定されている状況では困難であるが、解析対象試験の estimand を明確にした上で統合解析の計画や結果の解釈を行うことは重要である。解析方法の観点では、例えば、欠測データへの補完方法は、有効性の解析結果に影響し、試験結果を解釈する上でも重要であるが、MBMA 解析対象の試験間で欠測データに対して異なる補完方法が用いられている状況が想定される。欠測データに対して、各時点での評価項目の値を反復測定混合効果モデル (MMRM) により推定する、多重代入法により補完する、最終観測値、ベースライン値、最悪値等の単一の値を用いて補完 (それぞれ LOCF, BOCF, WOCF) する等、種々の解析アプローチがあり、MBMA 対象とする試験間で異なる方法で各時点での有効性が推定されている状況も多い。また、同一試験の報告であっても、主解析時点では欠測値を補完した解析結果が報告されているが、他の時点の結果は観測値の記述統計量での報告のみ、あるいはそもそも結果自体が論文中に提示されていない状況も考えられる。有効性の結果は estimand や補完方法の違いの影響を受け、特に長期試験等で中止割合が高い場合にはその影響が大きい。論文等の公表結果に基づく MBMA では通常患者個別デ

ータの入手ができないため、異なる方法で追加解析を実施することは通常不可能である。解析対象試験の **estimand** や補完方法が異なる場合、結果を解釈するにはその影響や **limitation** を考慮する必要がある。興味の対象となる **estimand** や補完方法（例えば MBMA の結果を新たな試験デザインの検討に活用する場合には、その試験で使用予定の補完方法）を用いて解析された報告のみを解析対象にする、あるいは方法ごとに複数の独立した MBMA を実施するアプローチも考えられるが、これらは本来 MBMA 解析に含めるべき試験の一部に絞った統合解析となり、バイアスの入った解析結果となる可能性に留意が必要である。すべての試験で記述統計量が入手できる場合は、記述統計量を使って MBMA 解析を実施することも考えられる。あるいは、異なる補完法で得られたすべてのデータを統合し、補完法を共変量としてモデルに含めて解析する方法も考えられるかもしれない。

また、試験間での評価時点が異なる状況も考えられ、各時点の解析結果にどの試験が寄与しているか確認することも重要である。論文中ではベースラインと主解析時点のみの報告であっても、実際には複数時点での解析が実施されている場合も考えられ、企業や規制当局のウェブサイトに掲載された原資料へアクセスできる場合、未公表の時点データを補完できる可能性がある。

## 2.3 MBMA モデルの構築

MBMA モデルの構築はリサーチクエスションの設定から始まり、解析方針の決定、探索的なデータ解析を経て、変量効果構造及び固定効果構造を段階的に検討する反復的なプロセスである。各段階においてモデル診断を行い、仮定や構造の妥当性を確認しながら最終モデルを構築する。[図 8](#)に全体のフローを示す。

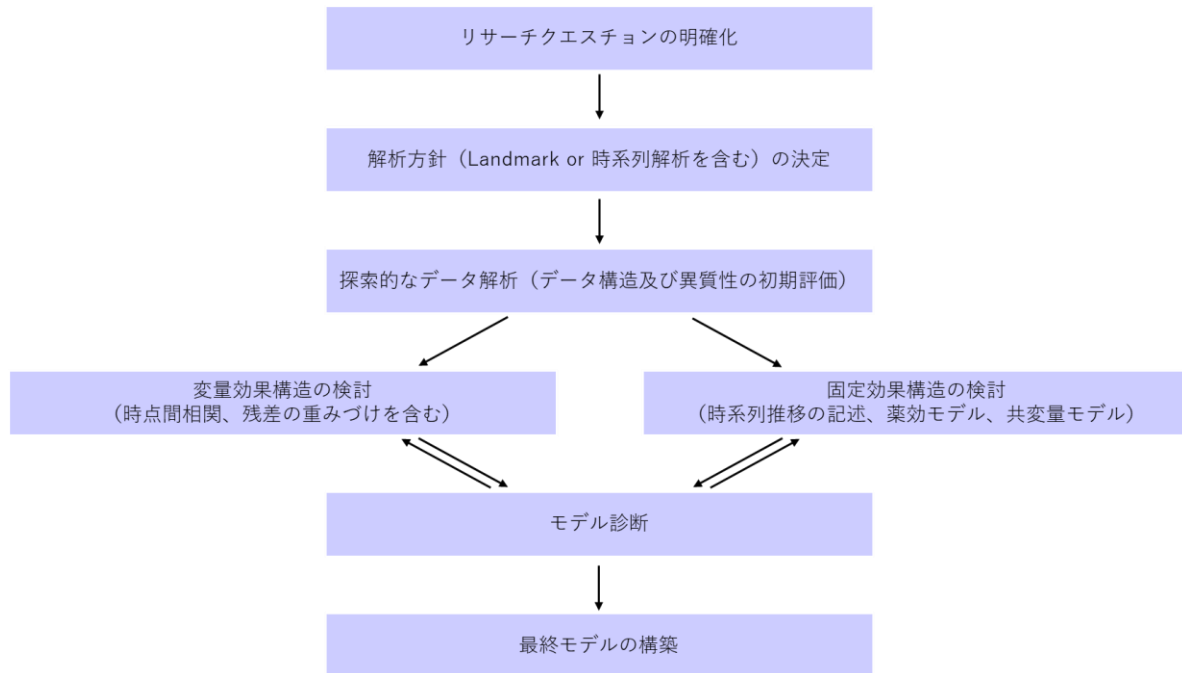


図 8 メタアナリシスのためのモデル構築のフローチャート

### 2.3.1 Landmark 解析と時系列解析のすみ分け

MBMA では、Landmark 解析と時系列 (Longitudinal) 解析のいずれを選択するかが重要であり、目的・データ特性・開発段階に応じて使い分ける必要がある。

Landmark 解析は、特定時点 (例：Week 12, End of Study) における治療効果を統合する枠組みである。評価時点が論文間で揃っている場合には実装が容易であり、以下に述べる時系列解析に比べて、必要な仮定が減る。一方で、時間推移の情報を捨てるため、作用発現の速さ・維持・減弱といった薬効プロファイルの差異が意思決定に重要な場面では情報が不足する。Landmark 解析の事例としては Shulgin らによる免疫チェックポイント阻害剤の用量と有害事象についての Landmark 解析[25]、整形外科手術後の静脈血栓塞栓症予防のための抗凝固剤の有効性と出血についての Landmark 解析が挙げられる[26]。

時系列解析は、文献に掲載されるグラフや複数時点の表データを用いて反応の時間経過をモデル化し、薬剤の反応プロファイル (効果の発現・消失時期、持続性など) を定量的に把握できる。これにより、短期データから長期成績を推定するなど、開発計画 (試験期間や評価時点の設定) に直接有用な予測を提供し得る。時系列解析の事例としては Maringwa らによる抗糖尿病薬と心拍数、収縮期血圧および拡張期血圧の時系列解析[27]、Zhong らによるフルチカゾンフランカルボン酸エステル及びフルチカゾンプロピオン酸エステルと呼吸機能の改善度の時系列解析[28]が挙げられる。Checchio らは乾癬の治療と Psoriasis Area and Severity Index (PASI) スコアの改善度について

ての Landmark 解析と時系列解析を実施しており[29]、時系列解析の利点の一つとして薬剤間の効果発現速度を比較できることと述べられている。

したがって、目的が「ある時点の比較・順位付け」であれば Landmark 解析を優先し、作用発現・維持・長期外挿が必要な場合には時系列解析を選択する。なお、時系列解析では同一試験内で複数時点が観測されることにより時点間相関が生じるため、相関構造の明示的な取り扱いが重要となる(2.3.3項参照)。相関を無視すると推定値の標準誤差の過小評価や、バイアスの発生、多時点を報告した試験への過剰な重み付けにつながる可能性がある。最終的には、研究課題(research question)に必要な解像度とデータの利用可能性のバランスを踏まえ、Landmark 解析と時系列解析のいずれを採用するか判断する。

### 2.3.2 変量効果構造の設定(試験間差, 群間差, 個体差, 残差)

MBMA では、試験間の違い(対象集団, 実施地域, 評価手法, 背景治療など)に起因する試験間差を変量効果として表現し、推定の不確実性と将来試験への外挿可能性を適切に反映することが重要である。Landmark 解析では、治療効果(例: プラセボとの差)に対する試験間分散を導入する変量効果モデルが基本となり、推定値に加えて予測区間を提示することで「将来試験での成績の幅」を表現できる。

次に、変量効果を試験間差・群間差・残差として分解する際には、利用可能なデータ量や情報の粒度(試験数, 観測時点数, 分散情報の有無)に応じて推定の安定性が左右される。特に試験間変動の推定は試験数に依存し、試験数が限られる場合には推定が不安定になり得るため、必要に応じて事前分布の活用も検討される。

時系列解析では、同一試験・同一治療群内で複数時点のデータを用いるため、時点間相関を考慮する必要がある。相関を無視すると推定値の標準誤差の過小評価や、バイアス、多時点を報告する試験が過度に影響するなどの問題が生じ得る。このため、相関構造(2.3.3項参照)を導入するとともに、必要に応じて試験間差に加えて、群間差を変量効果として入れるとよい[30]。このとき、群間差は試験間差の中にネストされるべきである。さらに、要約データを中心とする MBMA では、個体差は直接観測できないため、通常は試験・群水準に集約された変動として取り込まれる。一方、individual participant/patient data (IPD) が利用可能な場合には、患者レベルの個体差を階層モデルとして明示的に扱う拡張が可能であり、患者背景共変量の推定や患者レベルの予測の妥当性向上が期待できる。

### 2.3.3 時点間の相関の取り扱い

時系列解析では、同一試験・同一治療群における複数時点の要約値(平均値など)は互いに独立ではなく、時点間で相関することが一般的である。これは、同一集団(同一試験・同一群)に由来するデータを時系列に観測していることに加え、時間経過に伴う脱落や補完の影響により、各時点の要約値が関連し得るためである。

この時点間相関を適切に取り扱わず、残差を独立と仮定して解析すると、推定値の標準誤差の過小評価になり、バイアスや、多時点を報告した試験に過剰な重みがかかる可能性がある。

相関の取り扱いには、大きく分けて(1)階層変量効果による表現と(2)残差相関構造の導入の2つのアプローチがある。(1)では、試験レベルの変量効果に加えて治療群(arm)レベルの変量効果を導入し、arm レベルを試験レベルにネストさせることで、同一 arm 内の反復観測が共有す

るばらつきをモデル化し、結果として時点間相関を吸収する。(2)では、残差に対して Compound Symmetry (CS) や Autoregressive (AR(1))などの相関構造を仮定し、誤差共分散を明示的に表現する。例えば、R パッケージの nlme 等では CS や AR(1)を含む複数の相関構造が実装可能であり、解析環境に応じて現実的な選択肢として利用できる。

実務上は、まず構造モデル (時間推移モデル) の妥当性を確保したうえで、残差の診断を通じて相関構造の必要性を判断することが推奨される。具体的には、試験×治療群ごとに残差の時系列プロットを作成し、正負の残差が連続して出現する場合には、時点間相関が十分に説明できていない可能性が示唆される。この場合、初期段階では CS を導入し、改善が不十分な場合に AR(1)などより柔軟な構造へ拡張するとよい[5]。

### 2.3.4 残差の重みづけ

要約データを用いる MBMA では、同じエンドポイントであっても試験・群・時点ごとに推定精度が異なるため、各観測点の寄与を適切に反映する目的で残差 (観測誤差) を重みづけして推定することが重要である。特に時系列解析では、グラフや表として提示された時系列の要約データを解析対象とするため、精度の差異を考慮せず、単純に当てはめると推定の偏りや不確実性の過小評価につながり得る。

要約データで平均値が得られる場合、観測値  $Y_{ikt}$  (試験  $i$ , 群  $k$ , 時点  $t$ ) の推定誤差分散は通常、標準誤差 (SE) から近似できる。したがって、推定の基本は  $1/SE_{ikt}^2 = 1/(SD_{ikt}^2/N)$  に比例する重みづけ (逆分散重みづけ) を採用する。この場合、残差変動の標準偏差に相当する SD はデータから新たに推定せず、論文で報告された SE をそのまま誤差の大きさとして使用することがある。なお、前項の時点間の相関を含める場合には、その相関パラメータについては、推定している事例もある[27]。

一方で、最尤法に基づくモデルでは、残差変動の標準誤差 (又は標準偏差) を推定パラメータとして扱うことが一般的である。そのため、要約データの MBMA においても SE を固定値として扱うのではなく、誤差分散を  $SD^2/N$  の形で表現し、SD を推定するアプローチも用いられる。この方法ではサンプルサイズによる相対的な精度の差を反映しつつ、残差分散のスケールはデータから推定することとなる。これにより、解析に用いる要約データの精度のみならず、いわゆる残差 (実測値と予測値のズレ) も加味した SD の推定値を集約して得ることが可能となる。

公表要約データでは、時点別の平均は図表から取得できても、SD や SE の値を入手できない (報告されていない場合や figure が不明瞭でデジタイズが困難な場合等) ことが少なくない。この場合、重みづけを行うための情報が不十分となるため、以下の実務的対応を検討する。第一に、利用可能な範囲で SD や SE を収集し、欠落がある場合は SD の時間推移を可視化して補完の妥当性を検討する (例: 類似時点・類似試験からの補完, 時点間で SD が一定とする仮定など)。第二に、SD や SE の欠落が広範で補完の仮定が強くなる場合は、解析の頑健性を優先し、サンプルサイズに比例した重みづけなど、簡略化した重みづけを候補とする。ただしサンプルサイズでの重みづけは試験間・時点間で分散が同程度という仮定を含むため、適用可能性は疾患領域・評価尺度・試験設計を踏まえて慎重に判断する。また、分散の補完に関する方法をまとめた文献も報告されており[31], 参照されたい。

また、時系列データでは各論文で時点数が異なるため、SD や SE の欠落や時点数差が大きい場合には、Landmark 解析に落として比較時点を揃えることも、実務上の安全策となる。

更に、時系列解析では [2.3.3 項](#) に記載の通り残差の時点間の相関を考慮し、重みづけの前提（独立性）を診断することが推奨される。

### 2.3.5 共変量探索

MBMA における共変量探索は、試験間の異質性を、報告されている試験デザインや対象集団の特性の違いによって説明し、推定の一般化可能性や予測精度を高めることを目的として実施する。薬効の用量反応・時間反応を薬理学的に妥当なモデルで表現しつつ、試験間差の要因を共変量として取り込むことで、競合比較や試験デザイン検討などの意思決定に用いる定量的枠組みを得ることが可能となる。

一方で、要約データ中心の MBMA では、共変量は多くの場合「試験または治療群レベルの平均値（例：平均年齢、平均ベースライン重症度、女性の割合、地域割合など）」として扱われるため、一般的に各共変量の検出力は低いことが知られている[\[4\]](#)。集団レベルで得られた関連を個人レベルの効果として解釈すると **ecological bias** ([2.3.9 項](#)参照) が生じ得る点に留意が必要である。要約データの共変量関係は患者個人での関係性と一致しない可能性があるため、個別データで確認することが重要である。

共変量候補は、薬理学的妥当性及び試験設計上の交絡可能性を踏まえて事前に仮説として整理し、候補を限定した上で段階的に導入することが推奨される。

共変量の候補としては、疾患領域に依存するが、主に以下のものがある[\[7, 10\]](#)：

- 薬剤クラスとしての効果、及び個別薬剤の効果
- 用量及び投与レジメンの効果
- 試験デザインの影響（エンドポイント、比較対照、プラセボ/対照薬の成績、試験期間、反応評価時点など）
- 患者集団特性の影響
- 出版年（背景治療の変化を示す）

Landmark 解析では、共変量は主として単一時点の効果（例： $\Delta$ 、治療差、反応率など）に対して回帰として導入される。その際、共変量が報告されていない試験が混在しうること、また集団が似ていると共変量の値域の幅が狭いことに留意する。また、時系列解析では、Landmark 解析よりも共変量を当てはめ得るパラメータが増える可能性がある。

近年のレビューでは MBMA 研究の方法論・報告の質に改善余地があることが示されており、共変量探索においても解析方針（候補設定、採否基準、感度解析）を透明に記載することが重要である。

### 2.3.6 異質性の評価と対応

異質性（heterogeneity）は、患者背景・試験デザイン・評価法の差により、治療効果が試験ごとに系統的に異なる現象、つまり試験間変動を表す用語である。MBMA ではこの異質性をモデルに組み込むことが多いが、その程度を評価する統計量として様々なものがあり、Higgins らは  $I^2$  と  $H^2$  が有用であると述べている[\[32\]](#)。以下に異質性の評価に用いられる統計量を記載する。

Cochran's Q：固定効果の推定値からのズレの加重和。Q は自由度  $r-1$  の  $\chi^2$  分布に従う。問題は、

試験数が少ないと検出力が低く、多いとわずかな異質性でも有意になりやすくなり、メタアナリシス間で単純比較できない点である。

$$Q = \sum_{i=1}^r W_i (\hat{\theta}_i - \hat{\theta}_{FE})^2$$

$$W_i = \frac{1}{s_i^2}$$

$\hat{\theta}_{FE}$  : r 個の試験での固定効果,  $s_i^2$  :  $\hat{\theta}_i$  の標準誤差

$I^2$  : 試験間の効果量のばらつきの割合。  $I^2$  の値が大きいほど異質性が高い。

$$I^2 = 100\% \times \frac{(Q - (r - 1))}{Q}$$

$H^2$  : 試験間の効果量のばらつきの絶対的な度合いを示す。  $H^2$  の値が大きいほど異質性が高い。

$$H^2 = \frac{Q}{(r - 1)}$$

試験間分散 ( $\tau^2$ ) : 変量効果効果モデルでの解析から得られる試験間分散で、異質性の大きさを表すが、影響の大きさそのものは表さず、尺度依存でメタアナリシス間比較には向かない。

予測区間 : 試験間の異質性があると将来の試験に対する不確実性が懸念となる。予測区間は、推定精度と試験間変動の両方を含むため、将来の「真の効果 ( $\theta_i$ )」の範囲を示す。臨床開発では、将来実施する試験での期待される結果を知ることが重要な場面も多く、予測区間の算出は有用である。

近似的な  $100(1-\alpha)\%$  予測区間は以下で求めることができる。

$$\hat{\theta}_{RE} \pm t_{r-2}^{\alpha} \sqrt{\{\hat{\tau}^2 + \overline{SE}(\hat{\theta}_{FE})^2\}}$$

$\hat{\theta}_{RE}$  : 変量効果の推定値,  $\overline{SE}(\hat{\theta}_{FE})$  : 変量効果の標準誤差,  $\hat{\tau}^2$  : 試験間分散の推定値

### 2.3.7 パラメータ推定方法とソフトウェア

MBMA のパラメータ推定は、(1) 最尤推定 (maximum likelihood estimation) に基づく頻度論的アプローチ、(2) 事前分布を組み込むベイズ推定、のいずれでも実施できる[33]。

最尤推定は MBMA における代表的なパラメータ推定方法のひとつであり、観測データに対する尤度を最大化するパラメータ集合を同定することを目的とする。この推定はニュートン・ラフソン法や Expectation-Maximization (EM) 法によって行われる。最尤推定はデータ構造に偏りがなく、尤度関数が適切に定義できる場合に有効である。

ベイズ推定は従来の頻度論的手法に代わるアプローチとして近年 MBMA でも利用されている。ベイズ推定では、過去の臨床試験データ、専門家の知見、予備的研究などに基づいて事前分布を設定し、新たに観測されたデータによって事前分布を更新して事後分布を得る。この推定はマルコフ連鎖モンテカルロ (MCMC) 法によって行われ、Gibbs サンプルングや Metropolis-Hasting アルゴリズムなどが用いられる。ベイズ推定は特に試験数が限られておりデータがスパースな場合、試験間・群間・時点間の階層構造が複雑な場合、検証的試験で未実用量や未評価の時点など、直接検証されていない条件への予測や外挿を必要とする場合に適している。

Landmark 解析では、一般的な固定効果/変量効果モデルを基本として治療効果と試験間分散を推定する。一方で時系列解析では、Emax 等の非線形モデルや混合効果モデルの枠組みを用いて、時間推移 (onset/plateau/offset) を同時に推定し、必要に応じて時点間相関も考慮する。

実装ソフトウェアとしては、R を用いた汎用メタアナリシス・混合効果モデルに加え、非線形混合効果モデルを扱える NONMEM や、ベイズ推定を行う OpenBUGS 等が利用される。

### 2.3.8 モデル評価・妥当性確認

MBMA は、複数試験の要約データや異なる試験デザインを統合し、用量反応や時間推移を定量的に記述・予測することを目的とする。一方で、解析に用いるデータは文献由来であることが多く、解析方法や欠測処理、分散情報の入手状況が試験間で一致しないという構造的制約を有する。このため、MBMA ではモデル構築そのものと同程度に、モデル評価及び妥当性確認を通じて、推定結果が前提条件に過度に依存していないことを示すことが重要となる。

モデル評価の基本は、構築したモデルが観測データを適切に再現できているかを確認することである。具体的には、測定値と予測値の差を試験、治療群、用量、観測時点などで層別して確認し、特定の条件下で推定値に系統的な偏り (過大または過小推定) が生じていないかを評価する。具体的には、以下の方法でモデルの評価・妥当性の確認ができる。

**Visual predictive check (VPC)** : VPC は MBMA で広く受け入れられている代表的な診断手法の一つである。この方法では、推定されたモデルパラメータに基づいて多数のデータセットをシミュレーションし、シミュレーションされた分布のパーセンタイルを観測データと比較する。観測値は予測値の帯に重ね合わせられ、モデルが観察結果の中心的な傾向と変動性を十分に記述しているかどうかを視覚的に評価できる。VPC は異なる研究条件や患者集団における予測性能を評価するために、投与量、共変量、または時間によって層別化することができる。

**Residual Diagnostics (残差診断)** : 残差診断は予測値と観測値の差を分析することによって、モデルの性能を評価するためのもう 1 つの基本的なツールである。標準的なプロットとして、残差対時間、残差対予測値、及び分位数-分位数 (Q-Q) プロットがある。これらのプロットは、モデルの misspecification、想定した分布からの逸脱、不当分散性、残差の独立性の違反などの問題を示すパターンを識別するのに役立つ。これらの問題を早期に発見することで、モデルの改良が可能になり、全体的な適合度と予測性能が向上する。

**感度分析** : 感度分析は、基礎となる仮定、データ構造、またはモデリング戦略の変更に対して、モデルの推論がどの程度頑健であるかを評価する上で極めて重要な役割を果たす。この手法では、影響を及ぼす可能性のある研究の除外、共変量関係の修正、ベイズ推定における事前分布の調整、変量効果の構造の変更など、モデルの構成要素を系統的に変化させ、その結果生じる主要アウトカムへの影響を観察する。感度解析は、モデルによる意思決定が規制や臨床に影響を及ぼす場合

に特に重要である。

クロスバリデーション：クロスバリデーションは、MBMA モデルの一般化可能性と予測性能を評価するために使用される内部バリデーション手法である。一般的な手順には、leave-one-study-out (LOSO) や leave-one-arm-out (LOAO) があり、一部データを解析から除外し、その予測再現性を検証することで、モデルの妥当性と頑健性を評価する。評価指標には、二乗平均平方根誤差 (RMSE)、予測区間のカバー率、キャリブレーション・プロットなどが一般的に使用される。

外部バリデーション：外部バリデーションは、モデル構築には使用していない独立したデータセット、例えば新規臨床試験のデータや、あらかじめ除外しておいた層別サブグループのデータに対して、開発したモデルを適用することを指す。これらの外部データセットにおいてアウトカムを正確に予測できることは、モデルの妥当性を強く支持する証拠となり、規制当局への申請、前向き臨床試験シミュレーション、あるいは適応拡大の検討における信頼性を高める。

近年のレビューでは、MBMA 研究の方法論的・報告の質が十分でないことが示されており、透明性の高いモデル評価と報告の重要性が強調されている[11]。したがって、MBMA の結果を意思決定に活用するためには、モデル評価（適合度・予測・感度解析）を計画的に実施し、仮定・限界・不確実性を明示することが推奨される。

### 2.3.9 要約データと個別データの統合

MBMA は、複数の臨床試験から得られた要約データを統合し、治療効果や安全性をモデルベースで比較・評価する手法であり、試験間変動や共変量との関係を定量的に扱える点に強みがある。一方で、要約データは試験・群レベルに集約されており、集約レベルの共変量関係が患者レベルの関係を必ずしも反映しない。これに対し、IPD は、疾患進行や治療反応を詳細に記述できるが、取得にはコストや契約上の制約が伴い、すべての試験から入手できるとは限らない。このような背景から、文献等から得られる要約データと臨床試験から得られる個別データを統合して MBMA を実施することは、エビデンスを最大限活用し、より現実的な臨床試験シミュレーションを可能にする有望なアプローチと考えられている。

IPD と要約データを統合することで、要約データ単独の場合と比べて推定精度の向上や、生態学的推論の改善が報告されている。しかしその一方で、データ階層の異なる情報を同一モデルに組み込むことは統計的仮定を複雑化させ、バイアスを導入するリスクも伴う。特に、要約レベルで観測される時間相関や変動構造を個人レベルにそのまま当てはめると、ecological bias が生じる可能性がある。例えば、アルツハイマー病を対象とした MBMA では、ADAS-cog スコアの平均的な疾患進行はほぼ線形に見える一方、個人レベルでは明確な非線形性が存在することが示されている。よって、MBMA において IPD と要約データを組み合わせる解析を行うことは、モデル仮定に関する課題やモデル構築の複雑性が増すことから、実務上は必ずしも一般的ではない。

比較的単純な統合方法としては、IPD を試験・群レベルに要約し、要約データと併せて解析する二段階法がある。この方法は実装が容易である一方で、IPD が持つ個体間変動や共変量情報を要約の過程で失うため、IPD を直接利用する利点は限定的となる。より柔軟な統合手法としては、階層的モデリングアプローチが提案されている。この枠組みでは、IPD と要約データを異なる階層のデータとして扱い、個人レベル、試験・群レベルといった複数の変動構造を明示的にモデル化する。具体的には、IPD には個体レベルの尤度を、要約データには試験・群レベルの尤度をそれぞれ定義し、共通の構造パラメータを介して両者を結合する。さらに、Ravva らは非線形モデルにお

ける要約値の集計に伴うバイアスを軽減するため、個人レベルのモデルから試験・群レベルの平均構造を近似して導出したうえ、IPD 尤度と要約データ尤度を同時に最大化する手法を提示している[33]。

### 2.3.10 臨床試験のエンドポイントが連続変数でない場合

個別データのモデル解析では、解析対象となるエンドポイントのデータが連続変数でない場合であっても、リンク関数を導入すること（例：二値データはロジット等）や分布を検討すること（例：カウントデータはポアソン分布や負の二項分布等）で比較的容易に対応することが可能である。一方、MBMA で扱うような要約データのモデル解析では、以下に示すようなことについて留意する必要がある。

通常、たとえばエンドポイントがイベントあり・なしの二値データの場合、個別データの解析では 0/1 のデータをそのまま用いることとなるが、サマリーレベルの公表データの場合はそのイベントの発現割合が解析の対象データとなり、その割合についてロジット変換等を施して解析することが多い[34]。その場合も、重みとして二項分布の分散（すなわち  $p$  を割合、 $n$  を例数としたとき  $p(1-p)/n$ ）を用いる等の対応で、パラメータ推定に伴いうるバイアスを軽減することが可能となる。また、生存時間データの MBMA ではカプランマイヤー曲線をそのまま digitize し、指数関数やワイブル関数に当てはめて解析を行う事例もこれまでは多かったが、打ち切りが考慮されていないことが欠点だった。これを克服すべく、論文のカプランマイヤー曲線の下に記載されているイベント及び打ち切り情報からカプランマイヤー曲線、さらには個別データを再現する方法が報告されている[35]。

## 2.4 研究の質の担保

システマティックレビューやメタアナリシスを意思決定において有用な情報とするためには、適切な方法で実施、結果を報告することが重要であり、質の向上のためにいくつか指針が公表されている。PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020[22]は、システマティックレビュー中での報告内容の網羅性に関するガイドラインであり、レビューを行った目的と実施方法、得られた知見について透明で完全かつ正確な説明を提供するための指針を示している。システマティックレビューの報告で含めるべき項目を、システマティックレビューの計画、実施、報告という一連の方法に関するチェックリストとして明示している。チェックリストは7つの章立て（タイトル、抄録、緒言、方法、結果、考察、その他の情報）で構成され、計 27 項目が設定されている。PRISMA はシステマティックレビューやメタアナリシスの国際的規範と言えるものであり、これらの項目に従って解析を計画、報告することは、研究の質を担保する上で有用である。チェックリストは PRISMA の website (<https://www.prisma-statement.org/>) 上で公開されている。また、他にも研究の質の担保のために作成されたツールが報告されており、システマティックレビューの方法論的な質の評価のためのツールとして AMSTAR-2 (AMeasurement Tool to Access systematic Reviews)[36]があり、研究選択やデータ抽出、データの統合などの 11 項目から構成される。Yang らは、2024 年時点までに報告された 143 の MBMA 事例を調査した結果、PRISMA 2020 と AMSTAR 2 のチェックリストの各項目の遵守率が低く、質の改善の必要性を主張している[11]。一方これらのガイダンスは MBMA 向けに作成されたものではなく、特に PK や PD モデリングを含む MBMA に対しては、MBMA 特異的な方法論・報告チェ

ックリストの整備が急務であると主張している。例えば、MBMA 実施において考慮すべき実務的ポイントとして、従来型のメタアナリシス等と明確に区別するために MBMA を使用していることをタイトルに明示すること、また introduction 中に MBMA を採用する理由や研究課題の重要性を記載することを推奨している。方法に関しては MBMA における重要なステップ（データの点検、モデル構築、モデル評価、シミュレーション）を必ず含め、通常のシステマティックレビューの方法に則り潜在的なバイアスを軽減することの重要性も指摘している。また、MBMA の解析結果を PRISMA 2020 に従って明確に示し、結果の考察では文献レビューの limitation、モデルにおける仮定が結果に与える影響について議論することの必要性も指摘している。なお、医療・健康分野の研究報告の質と透明性を高めることを目的とした Equator (Enhancing the QUALity and Transparency Of health Research) network の website 上では、2026 年の公開を目指し STAR-MBMA (Standards for Transparent and Accurate Reporting of Model-Based Meta-Analyses) が作成段階であるとの情報があり [37]、近い将来 MBMA に関する報告指針が公開されることが期待される。

### 3 MBMA モデルの活用

本項目では既報のMBMAの利用例を参照して、MBMAモデルの医薬品開発への実践について、概要と個別事例を紹介する。

#### 3.1 モデル活用の概要

MBMAの医薬品開発への応用について、幾つかのレビュー論文でまとめられている[6, 7, 10, 11, 38]。Yang ZらはMBMAの活用について、2023年時点での計143報のMBMAの文献報告事例(医薬品開発のみならず、医薬品の適正使用に用いた事例も含む)から、対象となった疾患領域及び用途をまとめている[11]。疾患領域は多岐に渡っていたものの、比較的多く(10報以上)報告されたものとしては、神経系疾患(n=19, 13.2%)、内分泌・栄養・代謝疾患(n=17, 11.8%)、悪性新生物(n=16, 11.1%)、筋骨格系または結合組織疾患(n=12, 8.3%)、呼吸器系疾患(n=11, 7.6%)が挙げられた。また、用途に関しては医薬品開発への活用に限って見ると、臨床試験デザインの最適化が最も多く(17.5%)、次いで開発意思決定(12.6%)、探索及びトランスレーショナル研究(3.5%)となっていた。

これらの総論及び後に述べる個別事例を基に、MBMAの医薬品開発への活用例について、開発相ごとに表1にまとめた。

開発初期の探索研究では、MBMAは主に動物-ヒト間のトランスレーショナルな予測に用いられ、前例のある標的や作用機序を持つ化合物に対し、成功事例・失敗事例を含む既存の知見を統合し、ヒトでの有効用量域や目標曝露量を予測する用途がある。既存の知見に基づき、*in vitro*での効力で補正した臨床での曝露量と有効性の関係をモデル化することで、候補化合物の開発見込みを早期に順位付けし、Physiologically Based Pharmacokinetics (PBPK)や安全性のモデルと組み合わせ、研究段階の意思決定を支援する。

早期臨床開発(第1相~第2a相 [Proof of Concept : PoC] 試験)では、探索的な試験で得られた短期的なエンドポイントやバイオマーカーから長期の臨床エンドポイントを予測し、成功確率の予測に基づく相移行の意思決定や用量選択等の試験デザインを支援する。また、プラセボ反応や疾患進行の時間経過をMBMAモデルで記述し、情報量の高いミニマムな試験期間の提案や評価時点の設計にも利用される。

後期臨床開発(第2b相~第3相試験)では、自社開発品データ及び競合品を含む外部データを基にしたMBMAモデルからの予測により、第3相試験の成功確率と競合品とのベンチマーキングを行い、第3相試験実施のGo/No go判断のサポートを行う。また、標準治療によるエンドポイントの時系列な推移とバラつきを推定し、対照群の効果の事前情報として活用することで、サンプルサイズ、層別因子、主要評価時点、対照選択等の試験デザインを定量的にサポートできる。加えて、外部コントロールやヒストリカルコントロールを利用できる場面では、疾患経過、プラセボ効果、試験間差を明示的に扱えるMBMAの利用可能性が高い場面である。

さらに承認後は、利用可能な曝露、効果、安全性の関係を外部データも含めて統合し、初回承認時には検討していない用量や投与スケジュールでのベネフィット/リスクをシミュレーションして仮説を生成し、適用追加や用法追加等のアイデアを得る。さらに、特別集団や併用薬の影響を共変量として取り込み、モデルに基づく個別化投与(MIPD)へ橋渡しすることも可能となる。

表 1 開発相ごとのリサーチクエストに対する MBMA の活用例

開発相	リサーチクエスト	MBMA により提供されるエビデンス	代表的な事例
探索研究	開発候補品の選択, ヒト曝露量の目標値の設定	過去の開発化合物について, <i>in vitro</i> での効力で補正した曝露と臨床効果の化合物横断的な統合解析を実施し, 曝露反応関係をモデル化。候補化合物のデータを用いた予測により, 化合物間の比較と共に, ヒト有効用量域, 目標曝露量を定量化する。	<a href="#">13, 39, 40</a>
	動物モデル, <i>in vitro</i> 試験系のヒトへの外挿性	過去の開発化合物について, 動物種間 (または <i>in vitro</i> での効力) と臨床効果を統合する MBMA を実施し, 臨床で観察された反応性との整合性を評価する。	<a href="#">個別事例 1 41-44</a>
早期臨床開発	臨床試験の成功確率の予測 用量範囲の設定	バイオマーカーと有効性, 安全性エンドポイントをリンクさせるモデルを構築し, 短期試験のデータから長期の臨床アウトカムを予測することで, 臨床試験の成功確率の推定及び用量選択をサポートする。	<a href="#">個別事例 2 17, 45-47</a>
	PoC 試験のエンドポイント 評価期間の設定	時系列な MBMA モデルによる反応の立ち上がり, プラトー到達時期を推定し, 短期エンドポイントと長期エンドポイントの相関性を定量化することで, 評価時点設定の根拠 (どこまで評価すれば長期のアウトカムを精度よく予測できるか) とする。	<a href="#">15, 49-51</a>
	PoC 試験でのプラセボ効果 とその影響因子の予測	公表されたプラセボ対照試験を MBMA で統合し, プラセボ効果の大きさと変動要因を定量化する。対象集団の特性 (疾患ステージ, 前治療など) や試験デザインによる要因 (治療以外の介入, 支持療法など) を共変量として組み込み, プラセボ効果や疾患の自然経過への影響を定量化する。	<a href="#">29, 52-55</a>
後期臨床開発	競合品や標準治療と比較した第 3 相試験成功確率と試験デザインによる影響	自社品データと外部データの統合で構築した MBMA モデルによる, 臨床エンドポイントのシミュレーションを実施し, 開発化合物のベンチマーキングを行う。仮想第 3 相試験の臨床試験シミュレーションを実施し, 試験デザイン (比較対照, 対象集団, 評価時点) ごとの成功確率を定量的に示す。	<a href="#">個別事例 3 56-59</a>

開発相	リサーチクエスチョン	MBMA により提供されるエビデンス	代表的な事例
	第 3 相試験での主要な層別因子の特定	多様な背景因子，臨床的，地理的因子を含む試験の有効性アウトカムを統合する MBMA モデルを構築し，予後因子として重要な因子を共変量解析で同定する。共変量の分布の違いによる影響と層別化の必要性を，シミュレーションで定量化する。	<a href="#">29, 60-62</a>
	直接比較試験がない治療に対する優位性	プラセボ対照，または実薬対照のランダム化比較試験を対象に，モデルベースの Network Meta-Analysis (NMA) を実施し，第 3 相試験で得られた対象薬のデータから，複数の比較対照に対する仮想的な直接比較試験をシミュレーションする。	<a href="#">19, 63, 64</a>
承認後	レジメンの最適化や代替レジメンの提案	有効性と安全性の曝露または用量反応関係の MBMA モデルを構築し，代替用量，代替レジメンのベネフィット/リスクプロファイルをシミュレーションし，承認後の用量最適化に向けた仮説を生成する。	<a href="#">個別事例 4 65-67</a>

Upreti VV et al(2019)[\[7\]](#)より作成

### 3.2 個別事例-1 開発早期：適切なバイオマーカーの選定

- 概要：

本事例は、GABA<sub>A</sub>受容体作動薬であるベンゾジアゼピン系及び関連薬剤 8 成分を対象に、第 1 相試験データを用いてバイオマーカーの選択性と感度を定量的に比較したものである。GABA<sub>A</sub>受容体には複数の  $\alpha$  サブタイプが存在し、 $\alpha 1$  は鎮静・睡眠、 $\alpha 2/\alpha 3$  は抗不安作用に関連するが、 $\alpha 1$  選択的作動薬と非選択的作動薬の薬効・副作用プロファイルの差異を定量化する必要があった。

PubMed から選定した 163 報の文献データ（被験者総数 2,990 名）を用いて、SEM (saccadic eye movement), VAS (visual analogue scale of alert), DSST (digit symbol substitution task), CFFT (critical flicker fusion test) の 4 つのバイオマーカーと有害事象（傾眠、めまい、頭痛）を評価した。解析では、各薬剤の PK モデルを構築後、バイオマーカーごとにプラセボモデルを構築し、薬剤効果は濃度依存的な線形モデルで表現した。最終モデルを用いて各薬剤の効果をシミュレーションし、diazepam 10 mg を基準に正規化して比較するとともに、in vitro 受容体親和性比との相関を検証した。

その結果、SEM は薬剤の違いを最も鋭敏に反映し、 $\alpha 1$  選択的作動薬では非選択的作動薬と比較して SEM 応答が小さく、in vitro での受容体親和性比と良好な相関を示した。これにより SEM が  $\alpha 2/3$  サブタイプ活性（抗不安作用）を反映する指標であることが支持された。一方、VAS, DSST, CFFT は選択性による差が小さく、全般的な中枢抑制作用を反映する指標と位置づけられた。また、DSST 応答が傾眠・めまい発現率と最も高い相関を示した。

本事例は、従来混沌としていたバイオマーカーの特性を MBMA により定量的に整理した好例である。プラセボモデルを丁寧に構築する一方で薬剤効果モデルは簡略化するという戦略的アプローチにより、解析目的に焦点を絞った効率的な解析を実現し、初期臨床開発における適切なバイオマーカー選択の指針として実用的価値を提供している。

表 2 個別事例-1（開発早期：適切なバイオマーカーの選定）の概要

項目	内容
薬剤名	diazepam, lorazepam, temazepam, alprazolam, flunitrazepam, zolpidem, zopiclone, zaleplon
作用機序	GABA <sub>A</sub> 受容体のベンゾジアゼピン結合部位に作用し、GABA による抑制性シグナルを増強
疾患領域	中枢神経系（不眠症、抗不安、鎮静）
MBMA を適用した開発ステージ	第 1 相試験
MBMA の用途	PK/PD 関係の統合解析、バイオマーカー感度比較、薬効と副作用の関連性評価
ソース	Ren YP et al (2015)[12]

- MBMA を適用した背景、目的：

GABA<sub>A</sub>受容体には複数の  $\alpha$  サブタイプ ( $\alpha 1, \alpha 2, \alpha 3, \alpha 5$ ) があり、薬理作用（鎮静、抗不安、認知影響）と副作用（傾眠、めまい）に関係する。 $\alpha 1$  サブタイプは鎮静や睡眠、 $\alpha 2$  及び  $\alpha 3$  サブタイプは抗不安、 $\alpha 5$  サブタイプは健忘や記憶障害に関連することが知られている。本論文では、

特に鎮静・睡眠誘発作用 ( $\alpha_1$ ) と抗不安作用 ( $\alpha_2$  及び  $\alpha_3$ ) に着目し、これらを識別するバイオマーカーを明らかにすることが目的とされた。個別試験では用量・時間・評価指標が異なるため、MBMAにより PK/PD モデルを統合し、薬効・安全性の比較を可能にすることを目指した。

- **MBMA に使用したデータ :**

Pubmedにおいて、1964年から2013年4月までに公開された文献を対象として2回の調査が行われた。それぞれの調査から2840件、1275件の文献が抽出され、レビューの結果、非選択的作動薬である diazepam, lorazepam, temazepam, alprazolam, flunitrazepam 及び  $\alpha_1$  選択的作動薬である zolpidem, zopiclone, zaleplon について、163報が MBMA の対象とされた。文献調査から解析対象の試験選択までのフローチャートはソースの論文を参照。

被験者総数は2990名で、2237の要約データ（濃度データ：716、バイオマーカー：1304、有害事象：217）が解析対象とされた。バイオマーカーは SEM (saccadic eye movement), VAS (visual analogue scale or alert), DSST (digit symbol substitution task), CFFT (critical flicker fusion test) を対象とし、いずれもベースラインからの変化量が用いられた。有害事象は傾眠 (somnia), めまい (dizziness), 頭痛 (headache) が対象とされた。

- **解析の概要 :**

- **PK モデル :**

はじめに各薬剤の PK モデルを構築（一次吸収・消失を伴う 1~2 コンパートメントモデル）し、PK パラメータを推定した。続いて行われた PK/PD モデル構築の際には、これらの推定した PK パラメータは固定された。PK モデルにおいて、試験間変動は変量効果で表現された。

$$P_{stu} = P_{pop} \times e^{\eta}$$

- **PK/PD モデル :**

各バイオマーカーに対するプラセボモデルと薬剤モデルは別々に構築された。バイオマーカーについては、絶対値が得られないケースが存在したため、ベースラインからの変化量 (CFB) が用いられた。残差は比例誤差と加法誤差の組み合わせで表現され、各群の被験者数の平方根で除することで元データの不均衡に対処された。

$$Y = \hat{Y} \times \left( 1 + \frac{\varepsilon_{pro}}{\sqrt{n}} \right) + \frac{\varepsilon_{add}}{\sqrt{n}}$$

ここで、Y は実測値、 $\hat{Y}$  はモデルによる予測値、 $\varepsilon_{pro}$  は比例誤差、 $\varepsilon_{add}$  は加法誤差、n は各群の被験者数を示す。

プラセボモデルは、バイオマーカーごとに別々に構築された。SEM は投与後早期に急峻に低下し、その後緩やかに回復する傾向が見られることから、指数と線形の複合モデルが採用された。VAS は時間依存性が認められなかったことから定数モデルが、DSST は時間とともに単調な増加傾向が見られることから線形モデルが、CFFT は飽和型の時間経過を示すことから指数モデルがそ

れぞれ採用された。

$$\begin{aligned}
 EFF_{P-SEM} &= -(P_{MAX} \times e^{\eta_1}) \times (1 - e^{-(\beta \times e^{\eta_2}) \times t}) + (PSLP \times e^{\eta_3}) \times t \\
 EFF_{P-VAS} &= -(PCT + \eta_1) \\
 EFF_{P-DSST} &= (PCT + \eta_1) + (PSLP + \eta_2) \times t \\
 EFF_{P-CFFT} &= -(P_{MAX} + \eta_1) \times (1 - e^{-(\beta \times e^{\eta_2}) \times t})
 \end{aligned}$$

ここで、 $P_{MAX}$  はプラセボ投与における最大効果、 $\beta$  は指数モデルの速度定数、 $PSLP$  は線形モデルにおける傾き、 $PCT$  は定数を  $\eta_1$ ~ $\eta_3$  は各パラメータの試験間変動を表す。

薬剤の効果は、単純にプラセボ効果に加算されると仮定した。また、薬剤効果のモデルは、全ての薬剤及びバイオマーカーにおいて濃度に対する線形モデルとして構築された ( $E_{max}$  モデル等も検討されたが、線形モデルが最も良いと結論付けられた)。筆者らはこの理由について、文献から抽出した情報がサマリーレベルのものが多く、ばらつきに関する情報が少なかったことを挙げている。

$$\begin{aligned}
 Eff_{i,j} &= EFF_{P-i} + EFF_{D-i,j} \\
 Eff_{D-i,j} &= -(SLP_{i,j} \times e^{\eta_{i,j}}) \times CONC_j
 \end{aligned}$$

ここで、 $i$  はバイオマーカー、 $j$  は薬剤を示す。

薬剤効果のモデルにおけるパラメータ推定は、プラセボモデルによる推定値を固定して行われた。なお、別途プラセボモデルと薬剤モデルのパラメータを同時推定して、逐次推定における値と比較することで、解析の頑健性が確認されている。

最終的な PK/PD モデルを用いて、各薬剤について、単回投与後の最大濃度におけるプラセボ補正後の効果をバイオマーカーごとに 1000 回シミュレーションした。中央値及び 90% 予測区間を算出し、diazepam 10 mg を基準として正規化して薬剤ごとの各バイオマーカーにおける効果を比較した。加えて、文献に基づく *in vitro* の受容体親和性 ( $K_i$ ) から構成した Affinity 比 ( $\alpha_{2/3}/\alpha_1$ ) と、モデルに基づくバイオマーカーの SLP (傾き) 比 (例: SEM/DSST) の相関を検証した。

SEM は、薬剤の違いを最も鋭敏に反映する指標であった。特に、 $\alpha_1$  選択的作動薬 (zolpidem, zopiclone) では、非選択的作動薬と比較して SEM への効果が相対的に小さく、同一効果を得るためにより高い等価用量 (diazepam 10 mg が引き起こすバイオマーカー反応と同じ大きさの反応を示す用量) が必要であった。この傾向は、*in vitro* での受容体親和性 ( $\alpha_{2/3}$  と  $\alpha_1$  の比) との良好な相関として示され、SEM が  $\alpha_{2/3}$  サブタイプ活性を反映しやすいバイオマーカーであることを支持した。一方、VAS, DSST 及び CFFT は、 $\alpha_1$  選択的作動薬と非選択的作動薬の間で類似した反応を示し、等価用量も近い値となった。また、これらの指標同士や、*in vitro* 親和性比との間に明確な相関は認められなかった。以上より、VAS, DSST 及び CFFT は、受容体サブタイプ選択性よりも全般的な中枢抑制作用を反映する指標と位置づけられた。

探索的な検討として、PK/PD モデルが予測するバイオマーカー変化と選択された有害事象 (傾眠, めまい, 頭痛) 発現との関係を検討した。DSST 応答が傾眠・めまい発現率と最も高い相関を示した。

- **本事例の考察：**

MBMAにより、薬効評価におけるバイオマーカーの選択性が明確化された。すなわち、SEMは抗不安作用 ( $\alpha 2/3$  作動) に特異的な指標であり、鎮静・催眠作用 ( $\alpha 1$  作動) に対する効果は検出しにくい一方で、VAS/DSST/CFFTは全般的な中枢抑制の指標となることが示された。

従来、多くの情報が混在しており、古典的なベンゾジアゼピン系薬剤 ( $\alpha$  サブタイプ非特異的な作用を示す) ではDSSTが効果指標として知られていたり (68-71),  $\alpha 2$ ,  $\alpha 3$  作動に対する評価にはSEMが適しているという文献 (68) が存在したりしていた。本事例は、このような混沌とした状況において、MBMAが情報を統合し新たな知見を提供した事例と考えられる。また、実用的な観点から、論文では健康成人を対象とした第1相試験においてこれらのバイオマーカーを評価することで、その時点でのgo/no go判断に活用できることが示されている。

本事例では、プラセボ投与時の各バイオマーカーの時間推移に基づき、バイオマーカーごとにプラセボモデルが構築された。この点は、本MBMAの品質を支えた重要な要素と考えられる。一方で、薬剤の効果については、薬剤やバイオマーカーによらず一律で濃度に応じて線形に増加するモデルが採用された。本MBMAでは薬剤間の比較というよりも、作用機序別のバイオマーカーの反応特性を明確化することに主眼が置かれており、そのためプラセボモデルが丁寧に構築されたと考えられる。このように、必要な部分は詳細にモデル化する一方で、解析の目的と関係の薄い部分は大胆に簡略化することも、効率的に意図する情報を得るための解析においては重要であると考えられる。

### 3.3 個別事例-2 検証試験前：用量設定試験の代替

- 概要：

本事例では、RSウイルス（RSV）中和抗体である Clesrovimab（販売名：ENFLONSIA）を対象に、中和抗体力価（Serum Neutralizing Antibody：SNA）と臨床効果の関係を定量化し、臨床試験における用量選択や試験デザインの最適化に MBMA を活用した。文献情報と開発中抗体のデータを統合し、SNA を感染防御の主要な予測因子とするモデルを構築した。本 MBMA を活用して、第 I 相の RSV チャレンジ試験の SNA の結果を基に、第 II 相用量設定試験を実施せず、第 III 相試験の用量設定を支援した。本事例は、断片的な情報を統合し、ばらつきのあるデータから信頼性の高い予測を可能にする MBMA の有用性を示すものである。

表 3 個別事例-2（検証試験前：用量設定試験の代替）の概要

項目	内容
薬剤名	Clesrovimab／ENFLONSIA
作用機序	RSV 中和抗体
疾患領域	感染症
MBMA を適用した開発ステージ	第 1 相試験終了後
MBMA の用途	RSV に対する SNA と臨床効果の関係の定量化、臨床試験における用量選択や試験設計の最適化
ソース	Maas BM et al (2021)[16]

- MBMA を適用した背景、目的：

RSV 感染症は乳児における入院の主要原因であり、季節性に流行する。既存の予防手段である Palivizumab は複数回投与が必要で、かつ対象は高リスク乳児に限定されている。開発品である MK-1654（Clesrovimab）は高い RSV 中和活性と長い半減期を有し、単回投与でシーズン全体をカバーできる可能性がある。

従来の用量設定は動物モデルや経験則に依存しており、ヒトでの曝露量・感染リスクの時間変動（季節性）を十分に考慮できていない。また、抗体価（SNA）と臨床効果の関係は既存文献で示唆されているが、断片的であり、統合的な解析が可能となれば、新規の RSV に対するワクチン又は中和抗体の適切な用量設定まで達成することができるかもしれない。

これらの背景において、MBMA を用いて RSV に対する SNA と臨床効果の関係を定量化するとともに、臨床試験における適切な用量選択や試験デザインの最適化が試みられた。

- MBMA に使用したデータ：

キーワードに基づき、PubMed で 2016～2018 年に初回検索を実施し、2019 年に追加検索を行った。これらの検索から合計 5,891 件の記録が抽出され、重複を除いた 1,069 件について概要を確認した。そのうち、以下のいずれかの情報を提供する 362 件の文献を抽出した。

1. RSV 感染症に対する有効性
2. モノクローナル抗体（mAb）またはワクチン投与後の SNA 力価、もしくは SNA 力価を算出するために十分な情報

さらに精査した結果、48 試験が適格と判断され、厳格な無作為化比較試験（Tier 1）と高品質の

観察研究 (Tier 2) に分類された。最終解析には、文献検索時点で非公開だった開発中 mAb に関する 5 試験を追加し、合計 53 試験を含めた。検索キーワード、最終解析に含まれた試験、PRISMA フローチャート等の詳細は、ソース文献 (Supplementary Methods) を参照。

これら 53 試験から、試験デザイン、治療内容、SNA 値、臨床アウトカム、患者背景を抽出した。臨床アウトカムは図 9 の 5 つの重症度レベル (無症候性感染、症候性感染、下気道感染 [LRTI]、入院及び ICU 入室) に分類した。なお、軽症群にはより重症の患者が含まれる場合があるが、重症群には必ずしも軽症患者が含まれるわけではない。

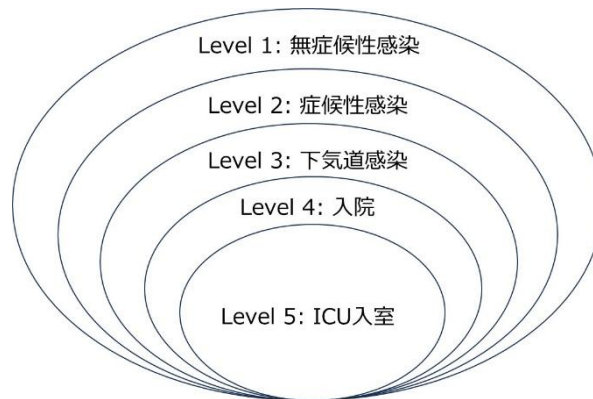


図 9 臨床アウトカム (重症度レベル) の分類

SNA の欠測値については、ワクチン試験では Day 29 の SNA 値から単純減衰モデルで補完した。mAb 試験では、成人における MK-1654 の PK データから構築した母集団 PK モデルを用い、報告された各試験の被験者背景と類似する仮想集団における MK-1654 の濃度推移をシミュレーションした。さらに、MK-1654 の濃度から線形関係で定義された SNA 値を、MK-1654 に対する各薬剤の相対強度 (非臨床試験に基づく) で補正し、欠測値を補完した。

• 解析の概要 :

<MBMA モデル>

本モデルは、「RSV 感染の発症リスクは、対象集団及び試験期間における SNA の時間推移のみに依存する」という基本仮定に基づいて構築した。年齢別の対象集団 (乳児、小児、成人及び高齢者の 4 区分) 及び重症度レベル (無症候性感染、症候性感染、LRTI、入院及び ICU 入室の 5 区分) ごとに、 $\log_2$ SNA 値と RSV 感染発症率の関係をシグモイド関数で記述した。発症率 (R) は、重症度レベル r、試験 j における試験期間 t での感染確率  $p(t)$  を積分して算出した。なお、i は投与群を示す。FOI(t) は RSV 流行強度 (季節性をガウス関数で補正)、 $N_{norm}(t)$  は観察期間中の被験者数補正を表す。

$$R_{r,i,j} = 100 \times \int_{t_{first}}^{t_{last}} p_{r,j}(t) \times FOI(t) \times N_{norm}(t) dt$$

$$p_{r,j}(t) = R_{max,r,j} \times e^{\log\left(\frac{R_{min,r,j}}{R_{max,r,j}}\right) \times Eff(\log_2 SNA(t))}$$

$$Eff(\log_2 SNA) = \frac{\log_2 SNA^\gamma}{\log_2 SNA^\gamma + IT50^\gamma}$$

$$FOI(t) = baseline + amp \times \frac{1}{\sigma \times \sqrt{2\pi}} \times e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

$$N_{norm} = \frac{N(t)}{N_{max}}$$

ここで、 $R_{max}$  は最大発症率（薬剤非投与時のシーズンピーク値）、 $R_{min}$  は最小発症率（薬効が最大限発揮された場合の値）を示す。 $\sigma$  はガウス分布の標準偏差、 $\mu$  は基準日（8月10日）から感染ピークまでの日数、 $amp$  はオフシーズンとピーク時の感染率差の大きさ、 $baseline$  はオフシーズンにおける感染率、 $N(t)$  は Day  $t$  に存在する被験者の割合、 $N_{max}$  は研究期間中の任意時点での最大被験者数である。

$R_{max}$  はロジット逆関数を用いて、前の重症度レベルの値を基準に次の重症度レベルの値を設定した。例えば、症候性感染における  $R_{max}$  は、無症候性感染における  $R_{max}$  に 0~1 の値（ロジット逆関数）を乗じて決定した。 $R_{min}$  は  $R_{max}$  に対する比（MinMaxRatio）として定義した。チャレンジ試験では、ウイルスを強制的に接種することや頻繁な検査により、フィールド試験と比較して感染率が高くなるため、別パラメータ（ $R_{max-chall}$ ,  $MinMaxRatio_{chall}$ ）を導入して補正した。

重みづけは RSV 感染発生率の精度に応じて行った。具体的には、重みは分散（または標準誤差の二乗）の  $p$  乗の逆数に比例し、パワー（ $p$ ）は 0.5 で固定した。

$$Weight \propto \frac{1}{VAR^{0.5}}$$

VAR は、モデルで推定した RSV 感染発生率（Poisson 分布を仮定）を、試験全体の曝露量（被験者数×RSV 感染シーズン数）で割った値とした。RSV 感染シーズン数は、試験期間中の被験者数の変動、RSV 流行強度（FOI）、検査率で補正した。

$$VAR = \frac{fitted}{total\ exposure}$$

Tier 2（観察研究）は Tier 1（RCT）より被験者数が多く試験期間も長いいため、そのまま解析すると重みが過大になる。そこで、Tier 2 の重みには  $VAR^{0.5}$  の代わりに  $VAR^{0.25}$  を用いた。

疾患重症度、患者集団、試験デザインごとのデータ量が十分でなく、個別にプロットしても曝露-反応関係の評価が困難であったため、 $V^2$ ACHER 法（スケーリングにより異なる条件のデータを統合評価する手法[72]）を用いて評価した。

モデルのバリデーションは、ラットチャレンジ試験、先行品目である REGN2222 の第 3 相試験、MK-1654 のヒトチャレンジ試験における実測値とモデル予測値を比較することで実施した。

<ラットチャレンジ試験>

Cotton rat に RB1 (MK-1654 の親抗体) 及び D25 (RSV の F タンパク質の site 0 に結合する mAb)

を、それぞれ 0.3～2.5 mg/kg の範囲で複数用量を単回静脈内投与 (Day 0) し、PK を評価した。その後、Day 1 に RSV A2 株を鼻腔内にチャレンジし、Day 4 に動物を安楽死させて鼻腔及び肺組織を採取し、RSV 量をプラークアッセイで測定した。これらの in-house データに加え、公開されている MEDI8897 のデータも統合し、MBMA モデルと同様のシグモイドモデルを用いて薬剤濃度とウイルス抑制効果の関係を定量化した。薬剤ごとの効果差は IT50 に対する共変量としてモデルに組み込み、試験ごとのウイルス力価の最大値及び最小値の違いもモデルで考慮した。

すべての抗体において、抗体濃度が高いほど肺の RSV 量は減少した。MBMA モデルは SNA と RSV 感染症発症の関係を定義しているため、健康成人における PK/SNA の関係を用いて抗体濃度を SNA 相当に換算し、さらに肺の RSV 量の最大値/最小値を  $R_{\max}/R_{\min}$  に対応させてスケールングした。変換後の曲線は、MBMA に基づくヒト乳児における SNA と RSV 感染症発症の関係を示す曲線と概ね一致することが確認された。

#### <先行品目である REGN2222 の第 3 相試験シミュレーションを通じたモデルバリデーション>

上記の MBMA モデルを用いて、REGN2222 (suptavumab) の臨床効果を予測した。仮想患者集団は、公開されている REGN2222 の第 3 相試験に関する論文の選択基準・除外基準に基づき、報告されている被験者背景と整合するように作成した。SNA の欠測値は、公開されている当該試験の PK データに基づき構築した母集団 PK モデルと、MK-1654 に対する REGN2222 の相対強度を用いて、PK/SNA の関係から補完した。

試験デザインに基づき、1,000 回の臨床試験シミュレーションを実施し、有効性はプラセボ群に対する実薬群の相対リスク (relative risk: RR) 低下率 ( $1-RR$ ) として算出した。1,000 回の試験結果から平均値及び 95%信頼区間 (logRR に対する正規分布を仮定し、 $\pm 1.96 \times$ 標準誤差) を求めた。

バリデーションは、RSVA 型に対する下気道感染発症の予防効果を対象に実施した。MBMA による予測値 (95%信頼区間) は、単回投与群で 70% (65%～75%)、2 回投与群で 81% (75%～85%) であった。一方、実測値 (95%信頼区間) は、単回投与群で 62% (-5%～86%)、2 回投与群で 61% (-7%～86%) であり、症例数が少ない影響もあり大きなばらつきが認められた。

#### <第 I 相 RSV チャレンジ試験>

健康成人男女 70 名 (18～55 歳) に、MK-1654 を 100, 200, 300, 900 mg またはプラセボで単回投与した (各群 14 名, Day 1)。Day 29 に RSV A Memphis 37b 株を鼻腔内にチャレンジし、その後 Day 31～39 まで 1 日 2 回、Day 40 に 1 回、鼻腔洗浄液のサンプルを収集した。ウイルス量は、鼻腔洗浄液を添加した細胞を培養し、RT-qPCR で定量した。併せて、チャレンジ前から Day 40 まで症状スコアを評価した。

SNA は用量依存的に上昇した。RSV 量の低下及び症候性感染の発症率は、実薬群で効果が認められた一方、200 mg 以上の用量では効果に差は見られなかった。モデルの予測性能については、無症候性感染 (重症度レベル 1) では、実測値とモデル予測範囲の重なりが良好であった。症候性感染 (重症度レベル 2) では、予測値は実測値より低い傾向を示したものの、全体として実測値はモデル予測範囲内に分布していた。これらの結果から、SNA が感染防御の予測因子となることが確認された。

#### <第 III 相試験を想定した臨床試験シミュレーション ー小児における MK-1654 事前投与による

## RSV 感染抑制ー>

米国の出生に関する疫学データ（CDC 提供「WONDER」データベース）を基に、早産児（妊娠 29～35 週）及び正期産乳児（妊娠 35 週以降）を想定した仮想集団を作成した。試験期間中の乳児の体重推移は、CDC の成長曲線に基づいて設定した。被験者の割り付けは MK-1654 群とプラセボ群で 2 対 1 とし、MK-1654 の用量は 10, 30, 50, 75, 100, 125 mg とした。

乳児の PK は、成人 PK データを基に構築した母集団 PK モデル（2 コンパートメントモデル、体重をクリアランス及び分布容積の共変量とする）を、体重に基づくアロメトリックスケールリングでスケールダウンしてシミュレーションした。SNA と PK は線形関係にあると仮定し、以下の式で表した：

$$SNA_{total} = SNA_{base} + Conc_{MK-1654} \times Slope$$

1,000 回のシミュレーションを実施し、仮想後期臨床試験における MK-1654 の RSV 感染（下気道感染）発症抑制効果を予測した。成長に伴う体重増加を考慮し、体重は試験期間を通じて変動する共変量としてモデルに組み込んだ。

シミュレーション結果では、MK-1654 単回投与において 10～50 mg の範囲で用量反応が認められ、75 mg 以上では予防効果が約 76% でプラトーに達し、効果は約 5 か月間持続した。この結果は、実際の第 2b/3 相試験（105 mg 単回投与 vs プラセボ）の計画に活用された。

### • 本事例の考察：

本事例は、乳児における RSV 感染症発症リスクを抑制するため、適切な MK-1654 の投与量設定に MBMA を活用したケースである。解析時点で In-house で実施した MK-1654 の有効性を検討する臨床試験は、健康成人を対象とした第 I 相チャレンジ試験のみであり、その結果を基に乳児を対象とした第 2/3 相試験の用量を設定した。このため、MK-1654 の開発パッケージにおいて当該 MBMA は非常に重要な役割を果たしたと考えられる。

本 MBMA は「RSV 感染のリスクは SNA によってのみ規定される」という仮定に基づいている。この仮定は、他のウイルス感染症におけるモデル解析でも採用されており [73]、mAb による感染防御は宿主の免疫機構（T 細胞や B 細胞を介するもの）と独立していることから、一定の妥当性があると考えられる。ウイルスのチャレンジ試験は倫理的な課題があり、特に小児や乳児を対象に実施することは困難であるため、成人チャレンジ試験の結果を基に乳児の用量設定を支援した点で、本 MBMA の意義は大きい。

論文の Discussion でも指摘されているように、断片的な情報を統合して考察できる点は MBMA の強みである。今回の MBMA では、サンプル数が少なく、試験ごとにデザインや検査法に差異があるものの、約 50 年にわたる情報を集約し、モデル内ではこれらを共変量ではなくばらつきとして表現しているため、予測には一定の信頼性がある。Cotton rat 試験では、肺における RSV 量を評価した結果、MBMA で推定されるよりも低い SNA で高い効果を示した（曲線は実測値が左にシフト）。このことは、動物モデルのみで用量設定を行った場合、効果を過大評価し、臨床試験で用量不足となるリスクを示唆している。また、REGN2222 のケースでは実測値に大きなばらつきが見られた。特に開発初期の小規模試験では、サンプルサイズが小さく、実測データのみでばらつきを含めた予測を行うことは困難である。このような場合でも、複数の情報源を統合し、差異やその理由を考察することで、より精度の高い予測が可能となる。

### 3.4 個別事例-3 検証試験前：Go/no-go 判断のための競合薬との比較

- 概要：

Fenebrutinib は、B 細胞受容体及び好中球・肥満細胞などの Fc 受容体経路を介した免疫活性化を抑制することを目的に設計された Bruton's tyrosine kinase (BTK) 阻害薬である。中等度及び重度の関節リウマチ (RA) 患者を対象とした第 2 相試験において fenebrutinib の有効性と安全性が評価された。本事例では、有効性評価項目である ACR20/50/70 反応率について、第 2 相試験データを使用した曝露-反応 (E-R) モデルと、RA の標準治療薬のデータを使用した MBMA を活用し、第 3 相移行の意思決定 (Go/No-Go) を支援した。第 2 相試験の最高用量であった fenebrutinib 200 mg BID の有効性は、標準治療薬である adalimumab 及び tofacitinib と概ね同程度であり、増量による上乗せ効果は限定的であることが MBMA と E-R 解析から示唆された。これらの知見は、fenebrutinib の RA 適応の開発を中止するという企業の意思決定を支持するものであった。

表 4 個別事例-3 (検証試験前：Go/no-go 判断のための競合薬との比較) の概要

項目	内容
薬剤名	fenebrutinib
作用機序	Bruton's tyrosine kinase (BTK) 阻害
疾患領域	関節リウマチ
MBMA を適用した開発ステージ	第 2 相試験終了後
MBMA の用途	第 3 相移行の意思決定 (Go/No-Go) サポート
ソース	Chan P et al (2022)[ <a href="#">10</a> ], Chan P et al (2020)[ <a href="#">74</a> ]

- MBMA を適用した背景、目的：

BTK 阻害薬の fenebrutinib は、RA 領域においては新規の作用機序を持つ薬剤であり、過去のメトトレキサート療法 (MTX) で効果不十分 (MTX-IR) 又は過去の抗 TNF 療法で効果不十分 (TNF-IR) の中等度及び重度の関節リウマチ (RA) 患者を対象とした第 2 相試験で fenebrutinib の有効性と安全性が評価された。第 2 相試験では、プラセボ、fenebrutinib 50 mg QD, 150 mg QD, 200 mg BID 及び adalimumab 40 mg 隔週投与の 5 群で試験が実施され、有効性評価項目は 12 週時点の ACR20/50/70 反応率であった。

RA 領域では、複数かつ異なる作用機序の治療薬が市販されており、新規作用機序を持つ fenebrutinib の他剤との差別化 (特定集団における有効性、効果発現速度など) が開発戦略の重要課題であった。RA 領域では多数の治療薬の膨大な臨床試験データが蓄積及び公開されており、これらのデータを活用しモデルを用いて統合する MBMA アプローチが有効と考えられた。

本事例では主に下記の目的で MBMA が活用された。

- 第 2 相試験内のプラセボ群及び adalimumab 群の ACR 反応率が外部データと整合しているかを検証し、第 2 相試験の結果の妥当性を評価する。

- MTX-IR や TNF-IR などの患者集団特性及び MTX 併用の有無が ACR 反応率に与える影響を推定する。さらに、次相試験の想定患者集団における fenebrutinib の有効性を予測し、標準治療薬 (adalimumab 及び tofacitinib) との比較を行う。

- MBMA に使用したデータ :

Fenebrutinib の開発企業の Genentech では、RA 領域における 2012 年までの臨床試験データを収集し MBMA データベースを構築していた[48]。このデータベースを更新する形で、PubMed/Cochrane Library/Embase を「rheumatoid arthritis」で体系的検索し、Cochrane Handbook のガイドラインに則った抽出手順で 2017 年までの RA 領域のランダム化比較試験のデータを収集した。データベースには、プラセボ投与を受けた 6567 例 (引用元 : 51 報), adalimumab 投与を受けた 1871 例 (引用元 : 16 報), tofacitinib 投与を受けた 1132 例 (引用元 : 10 報) のデータを含んでいた。Fenebrutinib の第 2 相試験で出られたデータ (578 例) についても、試験レベルで要約しデータベースに統合した。

- 解析の概要 :

3 つの有効性評価項目 (ACR20/50/70 反応率) について、各反応率の経時的変化を 1 つの非線形混合効果モデルで取り扱い、3 つの評価項目間の関連性を考慮した同時モデリングを実施した。これらの評価項目は患者レベルでは 2 値データであるが、MBMA では試験レベルの ACR20/50/70 反応率をロジット変換し連続変数として取り扱った。対象薬剤  $d$  の時点  $t$  における ACR 反応率 ( $k=ACR20, 50, 70$ )  $p_{k,d}(t)$  を下記の概念式で記述した。

$$\text{logit}\{p_{k,d}(t)\} = b_k + \eta_{k,trial} + \eta_{k,arm} + g_d(t) + f_d(dose) + cov + \varepsilon$$

ここで、 $b_k$  は各 ACR のベースライン、 $\eta_{k,trial}$  は試験間変動、 $\eta_{k,arm}$  は試験内群間変動、 $cov$  は共変量効果、 $\varepsilon$  は残差変動を示す。ベースラインは、 $ACR20 \geq ACR50 \geq ACR70$  の関係性が保たれるよう制約関数を組み込んだ。 $g_d(t)$  は時間効果であり、指数関数として最大効果へ漸近する形： $g_d(t) = 1 - \{e^{-(k+\eta_d)t}\}$  として記述された。ここで、 $k$  は速度定数、 $\eta_d$  は薬剤間変動を示す。 $f_d(dose)$  は薬剤別の用量-反応関数であり、薬剤ごとに線形、 $E_{max}$ 、あるいは定数式として記述された。用量は、異なる投与レジメンを比較可能とするため、週換算の用量を使用した。Fenebrutinib の MBMA のモデルの  $ED_{50}$  の推定は困難であったため、患者レベルのデータを使用した E-R 解析から得られた  $EC_{50}$  ( $AUC=2650 \text{ ng} \cdot \text{h/mL}$ ) に相当する用量を  $ED_{50}$  として使用した。共変量は forward addition 法で検討し、MTX-IR 及び TNF-IR の患者の割合、MTX 併用の影響が統計学的に有意な共変量として同定された。

Fenebrutinib の第 2 相試験内のプラセボ群及び adalimumab 群の ACR20/50/70 反応率の経時推移は MBMA から得られた推移と整合していた。本モデルを用いて、MTX-IR 及び TNF 阻害薬未投与集団に、fenebrutinib 200 mg BID、標準治療薬である adalimumab 及び tofacitinib を投与した際の 12 週までの ACR20/50/70 反応率の経時推移をシミュレーションした結果、fenebrutinib 200 mg BID を投与した際の 12 週時点の ACR20/50/70 反応率は adalimumab 及び tofacitinib と同程度であり、この結果は TNF-IR 集団でも同様であった。また、第 2 相試験内の曝露範囲で ACR20/50/70 反応率は、ほぼ  $E_{max}$  に到達していることが示唆され、仮に第 3 相試験で投与量を増量したとしても有効性の上昇幅は限定的と考えられた。以上の結果等から、有効性における標準治療薬との差別化は困難であることが示唆された。これらの知見は、fenebrutinib の RA 適応の開発を中止するという企業の意思決定を支持するものであった。

- **本事例の考察：**

本事例は、新規作用機序薬の開発相移行の意思決定（Go/No-Go）において MBMA による定量的な判断根拠を提供できた事例である。第一に、MBMA により、第 2 相試験内のプラセボ群及び adalimumab 群の ACR 反応率が、過去の多数の臨床試験から得られたヒストリカルデータと整合していることが確認され、単一試験結果の外的妥当性を検証する手段として有用であった。第二に、標準治療薬である adalimumab 及び tofacitinib を含む多剤の経時的データを統合し比較可能な枠組みを構築することで、fenebrutinib が有効性において他剤と同等であり、増量による有効性向上余地が小さいことを定量的に示し、他剤との差別化可能性を早期に判断できた点も意思決定に重要な役割を果たしたと考えられる。また、本事例に先立ち実施された別の MBMA（Wang Y et al (2016)[48]）において、12 週時点の ACR 反応率が中長期（24 週以降）時点の反応率を良好に予測する指標であることが、RA 領域の多数試験の統合解析により示されており、本事例では短期の 12 週までのデータを基に、24 週以降の長期有効性を見通しを得ることが可能であった。本事例から、MBMA は、新規薬剤の価値評価、競合環境でのポジショニング、及び開発継続判断において、不確実性を低減し、定量的根拠に基づく意思決定を実現する有用なツールであると考えられる。

### 3.5 個別事例-4 併用戦略：安全性プロファイルの評価

- 概要：

近年、がん領域において免疫チェックポイント阻害薬（Immune Checkpoint Inhibitors：ICI）に関連した数多くの臨床試験が実施され、幅広い患者集団から PD-1/PD-L1 阻害薬及び CTLA-4 阻害薬の安全性に関するデータが蓄積されている。本事例では、文献報告されている PD-1 阻害薬及び CTLA-4 阻害薬の単剤・併用療法の安全性データを MBMA により解析し、用量／曝露-有害事象関係と共変量の影響を定量化した。曝露推定には公開されている母集団 PK モデルを用い、定常状態平均濃度を算出して各薬剤の IC<sub>50</sub> で補正して使用した。解析の結果、PD-1 阻害薬単独では用量依存的な有害事象発現率の増加は認められなかったが、CTLA-4 阻害薬単独では用量・曝露依存的に有害事象が増加し、特に PD-1 阻害薬との併用療法では CTLA-4 阻害薬の曝露が低い場合でも Grade3/4 有害事象が高率で発生する相乗効果が示唆された。PD-L1 阻害薬のデータを追加して拡張したデータセットにおいても上記と同様な結果が得られた。これら MBMA で得られた知見は、durvalumab（PD-L1 阻害薬）と tremelimumab（CTLA-4 阻害薬）の併用療法の第 III 相試験における用量正当化を支持する根拠としても用いられた。

表 5 個別事例-4（併用戦略：安全性プロファイルの評価）の概要

項目	内容
薬剤名	nivolumab, pembrolizumab, ipilimumab, tremelimumab
作用機序	免疫チェックポイント阻害
疾患領域	がん
MBMA を適用した開発ステージ	承認後、第 III 相試験計画時
MBMA の用途	安全性の曝露反応関係の解析、用量妥当性のサポート
ソース	Chan P et al (2022)[10], Shulgin B et al (2020)[25]

- MBMA を適用した背景、目的：

ICI は幅広いがん種に対して効果を示す一方で、免疫関連有害事象（imAE）が発現しやすく、特に PD-1/PD-L1 阻害薬と CTLA-4 阻害薬の併用では有害事象が増強することが知られている。これまで個々の臨床試験やメタアナリシスで安全性は検討されてきたが、用量・曝露依存性に着目した定量的解析は系統的に行われてこなかった。本解析の目的は、ICI の異なる薬剤間・用量間の曝露を補正した上で、ICI の用量・曝露と重篤な治療関連有害事象の発現率との関係を MBMA で定量化することであった。

- MBMA に使用したデータ：

対象試験：PRISMA ガイドラインに従って、文献報告（学術論文及び会議抄録）から 2005～2018 年の ICI 安全性データを網羅的に検索し、基準に適合する 102 件の文献を選定した。これらには単剤及び併用療法のデータが含まれ、最終的に PD-1/CTLA-4 阻害薬治療の 153 コホート（試験群）、合計 21,305 名の患者データ（臨床試験 80 件相当）が解析に用いられた。Publication バイアスは funnel プロットと Egger 検定により評価し、効果量が対称性を保っていることを確認した。対象薬剤：PD-1 阻害薬として nivolumab, pembrolizumab, CTLA-4 阻害薬として ipilimumab,

tremelimumab を扱い、各単剤療法及び PD-1+CTLA-4 阻害薬併用療法のデータを統合解析した。PD-L1 阻害薬については、報告が限られていたため、本解析には含めていない。

有害事象：治療関連有害事象（trAE）の総発現率と Grade3/4 の trAE、及び imAE については胃腸障害、皮膚障害、肝障害、内分泌障害、肺障害の 5 つのサブグループ別の Grade3/4 発現率を解析した。Grade は CTCAE v4.03 基準に基づき分類した。

• 解析の概要：

ロジット変換した AE 発現率を目的変数とし、変量効果を考慮したメタ回帰解析を実施した。平均発生率が 5%未満の稀な AE に対しては、試験内変動に正規分布を仮定するのではなく、二項分布を仮定した一般化線形混合モデルを用いた。本解析で曝露の影響を検討するために、公開論文または FDA/EMA の公開資料から各薬剤の母集団 PK モデルの情報を入手し、定常状態の平均濃度をシミュレーションで算出した。得られた定常状態の平均濃度を各薬剤の *in vitro* 試験で報告されている親和性パラメータ（IC<sub>50</sub>）で補正したもの（C<sub>norm</sub>）を曝露パラメータとして使用し、異なる薬剤間で曝露を比較可能とした。下記の 4 つのモデルを候補として検討し、赤池情報量基準（AIC）などに基づき最適モデルを選択した。

(1) 加法モデル

$$\text{logit}(Pr_{AE}) = \beta_0 + \beta_1 \cdot C_{norm,CTLA4} + \beta_2 \cdot C_{norm,PD1}$$

(2) 加法モデル+CTLA-4 阻害薬曝露と PD-1 阻害薬曝露の交互作用項

$$\text{logit}(Pr_{AE}) = \beta_0 + \beta_1 \cdot C_{norm,CTLA4} + \beta_2 \cdot C_{norm,PD1} + \beta_3 \cdot C_{norm,CTLA4} \cdot C_{norm,PD1}$$

(3) 加法モデル+CTLA-4 阻害薬曝露曝露と PD-1 阻害薬併用有無（0/1）の交互作用項

$$\text{logit}(Pr_{AE}) = \beta_0 + \beta_1 \cdot C_{norm,CTLA4} + \beta_2 \cdot C_{norm,PD1} + \beta_3 \cdot C_{norm,CTLA4} \cdot Factor_{PD1}$$

(4) CTLA-4 阻害薬単剤曝露項のみ（PD-1 阻害薬単剤曝露項なし）+CTLA-4 阻害薬曝露と PD-1 阻害薬併用有無（0/1）の交互作用項

$$\text{logit}(Pr_{AE}) = \beta_0 + \beta_1 \cdot C_{norm,CTLA4} + \beta_3 \cdot C_{norm,CTLA4} \cdot Factor_{PD1}$$

ここで、 $\text{logit}(Pr_{AE})$ はロジット変換した AE 発現率、 $C_{norm,CTLA4}$ は CTLA-4 阻害薬の補正平均濃度、 $C_{norm,PD1}$ は PD-1 阻害薬の補正平均濃度、 $Factor_{PD1}$ は PD-1 阻害薬併用有無（0/1）を示す。また、試験レベルの患者背景因子として治療ライン（1st ライン vs 2nd ライン以降）、癌種（NSCLC vs メラノーマ vs その他）、化学療法併用の有無、ipilimumab vs tremelimumab, nivolumab vs pembrolizumab, 年齢中央値、性別構成、ベースライン PD-L1 陽性率などを逐次ステップワイズ法で共変量として探索し、統計的に有意な因子をモデルに組み込んだ。

解析結果から、Grade3/4 の trAE 発現率は CTLA-4 阻害薬の曝露増加に伴い上昇し、さらに PD-1 阻害薬併用（併用有り又は無し）で相乗的に発現率モデルが上昇するモデル(4)が最適と判断された。PD-1 阻害薬単剤の曝露量との関連は認められなかった。CTLA-4 阻害薬と PD-1 阻害薬との併用療法では、CTLA-4 阻害薬の曝露が比較的低い条件でも AE 発現率が著しく高い水準になる傾向が示された（[図 10](#)）。治療ラインについては、1st ラインの方が 2nd ライン以降よりも AE 発現率が高い傾向があり、また化学療法併用も AE リスク増加の共変量として同定された。一方、癌種別の差は認められず、試験間で癌種の異なる有害事象のデータを統合しても良いことが示唆された。また、胃腸障害及び肝障害に関連する imAE に関しても、上記と同様の結果が確認された。

さらに次の段階として、PD-L1 阻害薬の単剤投与及び併用投与を含む試験、ならびに 2018～2020 年に公表されたその他の ICI 試験のデータを追加して、MBMA データベースを拡張した。これにより、201 件の臨床試験・400 コホート・64,471 例から成る最大規模の ICI AE 関連のデータセットが構築された。この拡張データベースに、前述のモデルを適用したところ、同様な結果が得られた。すなわち、PD-1/PD-L1 阻害薬の単剤では安全性に対する曝露依存性は認められず、一方で PD-1+CTLA-4 又は PD-L1+CTLA-4 の併用療法では、CTLA-4 阻害薬単剤よりも高頻度の有害事象が観察された。注目すべきは、PD-L1 阻害薬の単剤における Grade 3/4 の trAE の発現率が PD-1 阻害薬の単剤よりも低いことであった。併用療法の状況でも同様に、CTLA-4 阻害薬+抗 PD-L1 抗体の組み合わせは、PD-1 阻害薬との併用よりも trAE 発現率が低い結果を示した。これらの結果は、既に報告されている論文報告と一致しており、PD-1 阻害薬は PD-L1 及び PD-L2 の両方との結合を阻害する一方、PD-L1 阻害薬は主として PD-L1 経路を介した末梢免疫寛容の解除に留まるため、毒性プロファイルに差が生じうるという解釈とも整合する。

これら MBMA の成果は、durvalumab (PD-L1 阻害薬) + tremelimumab (CTLA-4 阻害薬) の第 III 相試験における用量正当化を支持する根拠としても用いられた。

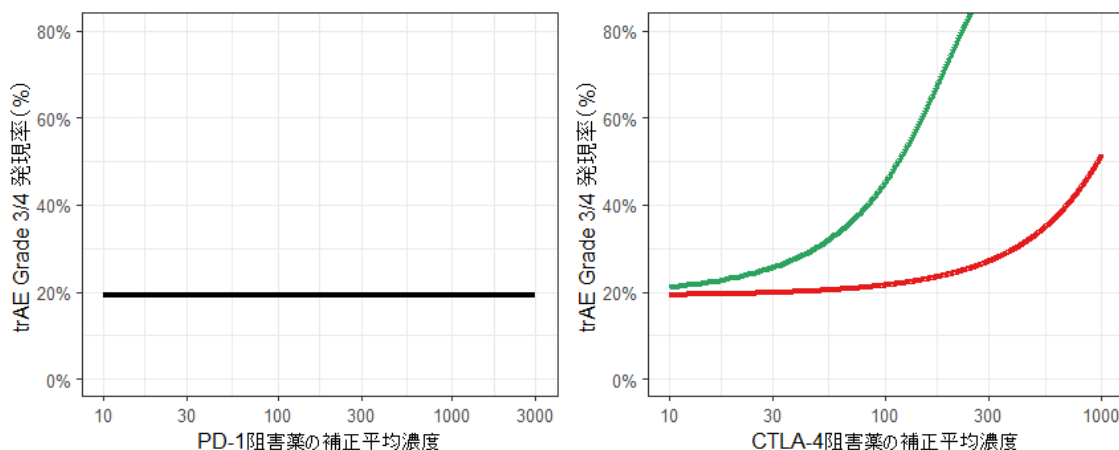


図 10 PD-1 阻害薬 (左図) 又は CTLA-4 阻害薬 (右図) の曝露量と Grade 3/4 の治療関連有害事象 (trAE) の発現率の関係

右図：緑色の曲線は CTCL-4 阻害薬と PD-1 阻害薬を併用した場合、赤色の曲線は CTCL-4 阻害薬単剤の場合の推定曲線を示す。

Shulgin B et al (2020)[25]より作成

• 本事例の考察：

メタアナリシスは一般的に有効性評価に用いられることが多いが、安全性評価にも適用可能であり、特に発現頻度の低い有害事象に対しては、複数試験のデータを統合したメタアナリシスが有用であるため、規制当局による医薬品安全性評価にも活用されている[75]。ICI を対象とした本事例では、学術論文、学会抄録などから得られたデータを統合したデータベースを構築し、薬剤の曝露量及び薬理学的効果を組み込んだ MBMA により、異なる治療法間の比較評価を実施した。定常状態の平均濃度などの各薬剤の曝露パラメータを公開済みの母集団 PK モデルからシミュレーションで推定し、さらに in vitro の IC<sub>50</sub> で補正した曝露指標に変換することで、異なる ICI 間の

薬理活性の差を補正した横断比較を可能にしており、これらの方法は異なる薬剤及び異なる用量間の曝露反応関係の検討に有用な方法と考えられる。

データベースは PRISMA に基づき体系的に構築され、まず 102 文献／153 コホート／21,305 例（約 80 試験相当）を解析し、出版バイアスは funnel プロットと Egger 検定で顕著な非対称性を認めないことが確認された。のちに PD-L1 阻害薬と 2018–2020 年の追加エビデンスを取り込み、201 試験／400 コホート／64,471 例まで拡張したデータベースでも拡張前の解析結果と同様の結論が再現されている。MBMA の結論としては、PD-1/PD-L1 阻害薬の単剤では安全性に対する曝露依存性は認められず、一方で PD-1+CTLA-4 又は PD-L1+CTLA-4 の併用療法では、CTLA-4 阻害薬の曝露が比較的低い条件でも AE 発現率が高い水準になる傾向が示された。また、併用では CTLA-4+PD-L1 の方が CTLA-4+PD-1 よりも AE 発現率が低い傾向が示された。これらの結果から、CTLA-4 阻害薬単剤又は PD-1/PD-L1 阻害薬との併用時に CTLA-4 阻害薬の曝露を臨床的に問題ない安全性の範囲にコントロールできるように用量を設定することの重要性、及び PD-1 阻害薬よりも PD-L1 阻害薬を併用パートナーに選ぶことが毒性低減に繋がる可能性が、本 MBMA によって定量的根拠を伴って示唆された。

## 4 まとめ

医薬品開発は不確実性が高く、膨大な開発コストと期間を要する一方で、成功確率は低水準にとどまっている。こうした課題への対応策として、MIDD は定量的な意思決定に基づいた開発の効率化を目指し、とりわけ MBMA は、公表データと自社データを統合したメタアナリシスモデルに、薬効・安全性の用量／曝露反応及び時間推移を組み込むことで、医薬品開発の各局面における定量的な意思決定に寄与する。MBMA は探索から市販後までの広範な開発段階に適用可能であり、競合比較、成功確率推定、試験デザインの最適化、承認後のレジメン検討に至る、意思決定に直結する予測を提供しうる。

MBMA の利用には主にデータベース構築・モデル構築・モデル活用の3段階が存在し、それぞれの段階において、MIDD で用いられる他の手法とは異なる留意点がある。MBMA モデルの構築と活用について、そのプロセスの概観と関連するトピックを [図 11](#) にまとめた。

データベース構築は MBMA の成否を左右する特に重要な過程であり、最も多くの労力を要するプロセスである。データ抽出時には研究の主題・PICO を明確化し、検索・選定・抽出を再現可能に設計し、重複や欠測、異なる estimand や欠測補完法の混在、評価時点の不一致等を体系的に扱うことが必要である。ここでは AI による文献探索、デジタルイズ、構造化の効率化が検討されているが、本報告書作成時の時点では未だ誤認識やセキュリティ・著作権の課題が残るため、専門家による検証を前提とした補助として位置付けとなっている。

MBMA で用いられるモデルでは、固定効果と多層の変量効果で真の治療効果とばらつきを表現し、試験レベル・群レベル・残差など、不確実性をそのソースに基づいて分離して推定する。解析方法には単一の時点を用いる Landmark 解析と、経時的な推移を取り扱う時系列解析が存在し、それぞれに利点・欠点がある。また必要に応じて、個別データと要約データの統合が可能であり、パラメータ推定精度と一般化可能性を高めうるが、階層の不整合から ecological bias (試験レベルのデータから見えた関係を患者レベルの関係と見間違う) を生みやすく、二段階要約や階層バイズなど設計と仮定の透明化が必要とされる。また、サンプルサイズによるデータの不確実性の影響を記述するために、症例数や SD/SE による残差の重みづけが行われるのも特徴的である。パラメータの推定では頻度論的手法とベイジック手法の双方が用いられ、モデル評価では VPC、残差診断、感度解析といった MIDD で用いられる他のモデル解析と共通したアプローチが用いられる。ここでシミュレーションによる予測の際には、信頼区間を提示して将来の試験結果の幅を明示することで、不確実性を考慮した意思決定を可能とする。

MBMA モデルの各医薬品開発段階での応用について、早期の探索段階では類効薬の用量・曝露反応を基に有効曝露目標を定め、早期臨床開発では短期指標から長期アウトカムの予測とそれに基づいた用量選択・試験デザインの最適化、他剤とのベンチマーキングを行い、早期の意思決定をサポートする。後期臨床開発では主に第 3 相移行の意思決定と試験デザインの最適化を行う。承認後は代替用量や併用療法のベネフィット／リスクを検討し、用法用量の変更・追加適応のアイデアを得ることが出来る。

このように MBMA は医薬品開発の広い範囲での活用が期待される手法である一方で、用いる手法の特徴からその他の MIDD で用いられる手法とは異なる専門性が求められ (AI を含むデータ収集・統計学的考慮・臨床試験デザイン最適化へのアイデア)、関係する専門家との緊密な協働が不可欠である。本報告書がこれらの議論を促し、MBMA を利用した医薬品開発の一層の浸透に貢献できることを期待する。

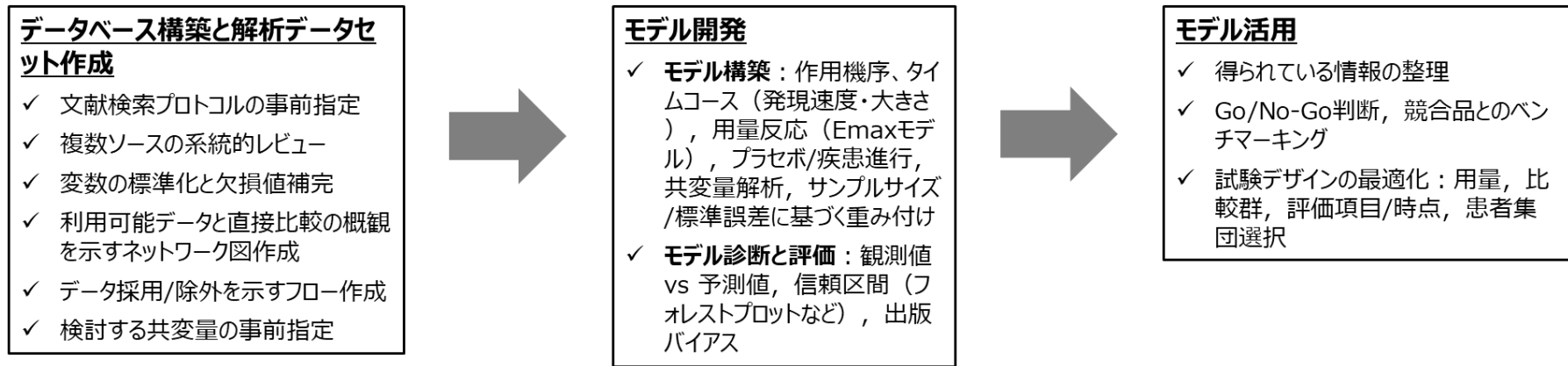


図 11 MBMA のプロセスの概観と各プロセスに関連する主なトピック

Chan P et al (2022)[10]より作成

## 5 引用文献

1. 日本製薬工業協会 医薬品評価委員会 データサイエンス部会 2022 年度/2023 年度継続タスクフォース 4 「国内承認品目での調査を通じた Model-Informed Drug Development の実状調査」. [https://www.jpma.or.jp/information/evaluation/results/allotment/DS\\_2022\\_mid.html](https://www.jpma.or.jp/information/evaluation/results/allotment/DS_2022_mid.html) [アクセス日: 12 Mar 2026].
2. 日本製薬工業協会 医薬品評価委員会 データサイエンス部会 2024 年度継続タスクフォース 4 「医薬品開発における Model Credibility 評価」. [https://www.jpma.or.jp/information/evaluation/results/allotment/DS\\_2024\\_mid.html](https://www.jpma.or.jp/information/evaluation/results/allotment/DS_2024_mid.html) [アクセス日: 12 Mar 2026].
3. EFPIA MID3 Workgroup. Good Practices in Model-Informed Drug Discovery and Development: Practice, Application, and Documentation. CPT Pharmacometrics Syst Pharmacol. 2016 Mar;5(3):93-122.
4. Boucher M, Bennetts M. The Many Flavors of Model-Based Meta-Analysis: Part I—Introduction and Landmark Data. CPT Pharmacometrics Syst Pharmacol. 2016;5(2):54-64.
5. Boucher M, Bennetts M. Many Flavors of Model-Based Meta-Analysis: Part II—Modeling Summary Level Longitudinal Responses. CPT Pharmacometrics Syst Pharmacol. 2018;7(5):288-297.
6. Mould DR. Model-based meta-analysis: an important tool for making quantitative decisions during drug development. Clin Pharmacol Ther. 2012;92(3):283-286.
7. Upreti VV, Venkatakrishnan K. Model-Based Meta-Analysis: Optimizing Research, Development, and Utilization of Therapeutics Using the Totality of Evidence. Clin Pharmacol Ther. 2019;106(5):981-992.
8. 厚生労働省, 2020 医薬品の曝露-反応解析ガイドライン. <https://www.pmda.go.jp/files/000235381.pdf> [アクセス日: 12 Mar 2026].
9. ICH, General Principles for Model-Informed Drug Development M15. <https://www.pmda.go.jp/files/000279201.pdf> [アクセス日: 12 Mar 2026].
10. Chan P, Peskov K, Song X. Applications of Model-Based Meta-Analysis in Drug Development. Pharm Res. 2022;39(8):1761-1777.
11. Yang Z, He H, Wang R, Liu D, Li G, Sun F. Application and Quality of Model-Based Meta-Analysis in Pharmaceutical Research: A Systematic Cross-Sectional Analysis and Practical Considerations. Clin Pharmacol Ther. 2024;116(2):397-407.
12. Ren YP, Xie RJ, Marshall S, Li L, Zhou TY, Lu W. Model-based meta-analysis of the effects of non-selective and  $\alpha$ 1-selective GABAA receptor agonists in healthy volunteers. Eur J Clin Pharmacol. 2015;71(10):1209-21.
13. Xu Y, Li YF, Zhang D, Dockendorf M, Tetteh E, Rizk ML, Grobler JA, Lai MT, Gobburu J, Ankrom W. Characterizing Class-Specific Exposure-Viral Load Suppression Response of HIV Antiretrovirals Using A Model-Based Meta-Analysis. Clin Transl Sci. 2016;9(4):192-200.
14. Zierhut ML, Chen Y, Pithavala YK, Nickens DJ, Valota O, Amantea MA. Clinical Trial Simulations From a Model-Based Meta-Analysis of Studies in Patients With Advanced Hepatocellular Carcinoma Receiving Antiangiogenic Therapy. CPT Pharmacometrics Syst Pharmacol. 2016;5(5):274-82.
15. Demin I, Hamrén B, Luttringer O, Pillai G, Jung T. Longitudinal model-based meta-analysis in

- rheumatoid arthritis: an application toward model-based drug development. *Clin Pharmacol Ther.* 2012;92(3):352-9.
16. Maas BM, Lommerse J, Plock N, Railkar RA, Cheung SYA, Caro L, et al. Forward and reverse translational approaches to predict efficacy of neutralizing respiratory syncytial virus (RSV) antibody prophylaxis. *EBioMedicine.* 2021;73:103651.
  17. Eudy-Byrne RJ, Gillespie W, Riggs MM, Gastonguay MR. A model of fracture risk used to examine the link between bone mineral density and the impact of different therapeutic mechanisms on fracture outcomes in patients with osteoporosis. *J Pharmacokinet Pharmacodyn.* 2017;44(6):599-609.
  18. 日本製薬工業協会 医薬品評価委員会 データサイエンス部会「ネットワークメタアナリシスの概要および留意事項 — ISPOR レポート(2011)を参考に」 .  
[https://www.jpma.or.jp/information/evaluation/results/allotment/lofurc000000b8vb-att/nma\\_20191110.pdf](https://www.jpma.or.jp/information/evaluation/results/allotment/lofurc000000b8vb-att/nma_20191110.pdf) [アクセス日: 12 Mar 2026].
  19. Mawdsley D, Bennetts M, Dias S, Boucher M, Welton NJ. Model-Based Network Meta-Analysis: A Framework for Evidence Synthesis of Clinical Trial Data. *CPT Pharmacometrics Syst Pharmacol.* 2016;5:393-401.
  20. Simeoni M, Chan P, Hang Y, Jreich R, Qiu J, Rigaux C et al. MBMA SubSIG: a collaborative example for promoting model-based meta-analysis and its application in drug development. *PAGE 30 (2022) Abstr 10224* <http://www.page-meeting.org/?abstract=10224> [アクセス日: 22 Apr 2026].
  21. 荒野俊平, 乙黒俊也, 角重喜, Model-based meta-analysis を用いた経時の試験結果の統合, SAS ユーザー総会 (2015) [https://www.sas.com/content/dam/SAS/ja\\_jp/doc/event/sas-user-groups/usergroups2015-b-07.pdf](https://www.sas.com/content/dam/SAS/ja_jp/doc/event/sas-user-groups/usergroups2015-b-07.pdf) [アクセス日: 12 Mar 2026]
  22. PRISMA. PRISMA 2020 flow diagram for new systematic reviews which included searches of databases, registers and other sources. <https://www.prisma-statement.org/prisma-2020-flow-diagram> [アクセス日: 12 Mar 2026]
  23. Cochrane, *Cochrane Handbook for Systematic Reviews of Interventions* <https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current> [アクセス日: 12 Mar 2026]
  24. McLellan J, Perera R. Restricted meta-analyses versus full meta-analyses: threshold number of studies based on study sample size *BMJ Evidence-Based Medicine* 2018;23:A11.
  25. Shulgin B, Kosinsky Y, Omelchenko A, Chu L, Mugundu G, Aksenov S et al. Dose dependence of treatment-related adverse events for immune checkpoint inhibitor therapies: a model-based meta-analysis. *Oncoimmunology.* 2020;9(1):1748982.
  26. Mandema JW, Boyd RA, DiCarlo LA. Therapeutic index of anticoagulants for prevention of venous thromboembolism following orthopedic surgery: a dose-response meta-analysis. *Clin Pharmacol Ther.* 2011;90(6):820-7.
  27. Maringwa J, Sardu ML, Hang Y, Czerniak R, Vishnubhotla M, Vakilynejad M et al. Characterizing Effects of Antidiabetic Drugs on Heart Rate, Systolic and Diastolic Blood Pressure. *Clin Pharmacol Ther.* 2021;109(6):1583-1592.
  28. Zhong Y, Li L, Chen R, Zheng Q. Quantitative comparison of dose-effect and time-course of fluticasone furoate and fluticasone propionate in adult and adolescent patients with persistent asthma:

- A systematic review and meta-analysis. *Respirology*. 2022;27(3):194-201.
29. Checchio T, Ahadieh S, Gupta P, Mandema J, Puig L, Wolk R, et al. Quantitative Evaluations of Time-Course and Treatment Effects of Systemic Agents for Psoriasis: A Model-Based Meta-Analysis. *Clin Pharmacol Ther*. 2017;102(6):1006-1016.
  30. Ahn JE, French JL. Longitudinal aggregate data model-based meta-analysis with NONMEM: approaches to handling within treatment arm correlation. *J Pharmacokinet Pharmacodyn*. 2010;37(2):179-201.
  31. Wiebe N, Vandermeer B, Platt RW, Klassen TP, Moher D, Barrowman NJ. A systematic review identifies a lack of standardization in methods for handling missing variance data. *J Clin Epidemiol*. 2006;59(4):342-53.
  32. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21(11):1539-58.
  33. Ravva P, Karlsson MO, French JL. A linearization approach for the model-based analysis of combined aggregate and individual patient data. *Stat Med*. 2014;33(9):1460-76.
  34. Turner DC, Wada R, Zhou H, Wang X, de Greef R, Valiathan C, Zhang L, Zhang N, Kuchimanchi M, Chen TT, Ballas M, Visser SAG. Model-based meta-analysis of non-small cell lung cancer with standard of care PD-1 inhibitors and chemotherapy for early development decision making. *CPT Pharmacometrics Syst Pharmacol*. 2023;12(11):1751-1763.
  35. Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Med Res Methodol*. 2012;12:9.
  36. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. 2017;358:j4008.
  37. Equator network <https://www.equator-network.org/library/reporting-guidelines-under-development/reporting-guidelines-under-development-for-systematic-reviews/#MBMA> [アクセス日: 12 Mar 2026].
  38. Sukumaran B, Xavier RM, Aswathy VS, Sharumathi SM, Rajamohamed H, Arun KP, et al. Beyond the Forest Plot: Redefining Model-Based Meta-Analysis as a Quantitative Engine for Model-Informed Drug Development. *Drug Dev Res*. 2026;87(1):e70227.
  39. Berg A, Clary J, Hanna D, Nuermberger E, Lenaerts A, Ammerman N et al. Model-Based Meta-Analysis of Relapsing Mouse Model Studies from the Critical Path to Tuberculosis Drug Regimens Initiative Database. *Antimicrob Agents Chemother*. 2022;66(3):e01793-21.
  40. Sandra L, T'jollyn H, Vermeulen A, Ackaert O, Perez-Ruixo JJ. Model-based meta-analysis to quantify the effects of short interfering RNA therapeutics on hepatitis B surface antigen turnover in hepatitis B-infected mice. *CPT Pharmacometrics Syst Pharmacol*. 2024;13:729-742.
  41. Niva C, Parkinson J, Olsson F, van Schaick E, Lundkvist J, Visser SA. Has inhibition of A $\beta$  production adequately been tested as therapeutic approach in mild AD? A model-based meta-analysis of  $\gamma$ -secretase inhibitor data. *Eur J Clin Pharmacol*. 2013;69(6):1247-60.
  42. Caruso A, Fűth M, Alvarez-Sánchez R, Belli S, Diack C, Maass KF et al. Ocular Half-Life of

- Intravitreal Biologics in Humans and Other Species: Meta-Analysis and Model-Based Prediction. *Mol Pharm.* 2020;17:695-709.
43. Kandala B, Plock N, Chawla A, Largajolli A, Robey S, Watson K et al. Accelerating model-informed decisions for COVID-19 vaccine candidates using a model-based meta-analysis approach. *eBioMedicine.* 2022;84:104264.
  44. Panday SK, Lang BJ, Kapitanov GI, Subramanian K, Klopp-Schulze L, Venkatakrishnan K, Zutshi A, Alnaif AE. A Translational Model-Based Meta-Analysis to Predict Tremor Incidence Associated with Serotonin Reuptake Transporter Inhibition. *Clin Pharmacol Ther.* 2025;118(3):588-592.
  45. Gibbs JP, Fredrickson J, Barbee T, Correa I, Smith B, Lin SL et al. Quantitative model of the relationship between dipeptidyl peptidase-4 (DPP-4) inhibition and response: meta-analysis of alogliptin, saxagliptin, sitagliptin, and vildagliptin efficacy results. *J Clin Pharmacol.* 2012;52(10):1494-505.
  46. Yao X, Zhou J, Song L, Ren Y, Hu P, Liu D. A model-based meta analysis study of sodium glucose co-transporter-2 inhibitors. *CPT Pharmacometrics Syst Pharmacol.* 2023;12:487-499.
  47. Bachhav SS, Ponce-Bobadilla AV, Clausznitzer D, Stodtmann S, Xiong H. Use of Model-Based Meta-Analysis to Inform the Design of Early Clinical Trials of Anti-Amyloid Beta Therapies in Alzheimer's Disease. *CPT Pharmacometrics Syst Pharmacol.* 2025;14:1191-1200.
  48. Wang Y, Zhu R, Xiao J, Davis JC Jr, Mandema JW, Jin JY et al. Short-Term Efficacy Reliably Predicts Long-Term Clinical Benefit in Rheumatoid Arthritis Clinical Trials as Demonstrated by Model-Based Meta-Analysis. *J Clin Pharmacol.* 2016;56(7):835-844.
  49. Plock N, Bax L, Lee D, DeManno D, Lahu G, Pfister M. Exploratory Literature Meta-Analysis to Characterize the Relationship Between Early and Longer Term Body Weight Loss for Antiobesity Compounds. *J Clin Pharmacol.* 2017;57(1):52-63.
  50. Leil TA, Lu Y, Bouillon-Pichault M, Wong R, Nowak M. Model-Based Meta-Analysis Compares DAS28 Rheumatoid Arthritis Treatment Effects and Suggests an Expedited Trial Design for Early Clinical Development. *Clin Pharmacol Ther.* 2021;109(2):517-527.
  51. Franzese RC, Qin L, Fu S, Rich B, Zografos E, Zierhut ML et al. Model-Based Meta-Analysis of Objective Response Rate and Survival Endpoints to Compare PD-1 and PD-L1 Treatment Outcomes in Non-Small Cell Lung Cancer. *CPT Pharmacometrics Syst Pharmacol.* 2026;15:e70196.
  52. Wu J, Xu L, Lv Y, Dong L, Zheng Q, Li L. Quantitative analysis of efficacy and associated factors of calcium intake on bone mineral density in postmenopausal women. *Osteoporos Int.* 2017;28(6):2003-2010.
  53. Li L, Xu L, Wu J, Dong L, Lv Y, Zheng Q. Quantitative analysis of placebo response and factors associated with menopausal hot flashes. *Menopause.* 2017;24(8):932-937.
  54. Arshad U, Rahman F, Hanan N, Chen C. Longitudinal Meta-Analysis of Historical Parkinson's Disease Trials to Inform Future Trial Design. *Movement Disorders.* 2023;38(9):1716-1727.
  55. Serrano JC, Maringwa J, Straetemans R, Willems W, Liva SG, Verhoeven J et al. A Model-Based Meta-Analysis Framework Quantifying Drivers of Placebo Response in Atopic Dermatitis Trials. *CPT Pharmacometrics Syst Pharmacol.* 2026;15(2):e70150.
  56. Chan P, Bax L, Chen C, Zhang N, Huang SP, Soares H et al. Model-based Meta-Analysis on the

- Efficacy of Pharmacological Treatments for Idiopathic Pulmonary Fibrosis. *CPT Pharmacometrics Syst Pharmacol.* 2017;6(10):695-704.
57. Llanos-Paez C, Ambery C, Yang S, Beerah M, Plan EL, Karlsson MO. Joint longitudinal model-based meta-analysis of FEV1 and exacerbation rate in randomized COPD trials. *J Pharmacokinet Pharmacodyn.* 2023;50:297-314.
  58. Yang L, Llanos-Paez C, Yang S, Ambery C, Berges A, Kjellsson MC et al. A Combined Model-Based Meta-Analysis of Aggregated and Individual FEV1 Data From Randomized COPD Trials. *CPT Pharmacometrics Syst Pharmacol.* 2026;15:e70059.
  59. Maloney A, Dua P, Ahmed GF. Comparative Effectiveness of Bimekizumab in Psoriatic Arthritis: A Model-Based Meta-Analysis of American College of Rheumatology Response Criteria. *Clin Pharmacol Ther.* 2024;115(5):1007-14.
  60. Korell J, Martin SW, Karlsson MO, Ribbing J. A model-based longitudinal meta-analysis of FEV1 in randomized COPD trials. *Clin Pharmacol Ther.* 2016;99(3):315-24.
  61. Slejko JF, Willke RJ, Ribbing J, Milligan P. Translating Pharmacometrics to a Pharmacoeconomic Model of COPD. *Value Health.* 2016;19(8):1026-1032.
  62. Yang S, Simeoni M, Beerah M. Longitudinal Model-Based Meta-Analysis of Lung Function Response to Support Phase III Study Design in Chinese Patients with Asthma. *Clin Pharmacol Ther.* 2022;111(6):1286-95.
  63. Yamamoto S, Horita N, Hara J, Sasamoto M, Kanemitsu Y, Hara Y et al. Benefit-Risk Profile of P2X3 Receptor Antagonists for Treatment of Chronic Cough: Dose-Response Model-Based Network Meta-Analysis. *CHEST.* 2024;166(5):1124-1140.
  64. Zivadnov R, Keenan AJ, Le HH, Al-Thiyaty M, Gandhi K, Zierhut ML et al. Brain volume loss in relapsing multiple sclerosis: indirect treatment comparisons of available disease-modifying therapies. *BMC Neurol.* 2024;24:378.
  65. Yoshioka H, Sato H, Hatakeyama H, Hisaka A. Model-based meta-analysis to evaluate optimal doses of direct oral factor Xa inhibitors in atrial fibrillation patients. *Blood Adv.* 2018;2(10):1066-1075.
  66. Asimwe IG, Chtiba N, Mouksassi S, Pillai G, Peter RM, Yuen E et al. Postmarketing Assessment of Antibody-Drug Conjugates: Proof-of-Concept Using Model-Based Meta-Analysis and a Clinical Utility Index Approach. *CPT Pharmacometrics Syst Pharmacol.* 2025;14:1810-1822.
  67. Chen J, Wada R, Zhang N, Graupner V, Morris S, Hu Y et al. Model-Based Meta-Analysis of Overall Survival in Vulnerable Platinum-Ineligible NSCLC Populations. *CPT Pharmacometrics Syst Pharmacol.* 2026;15:e70197.
  68. Chen X, de Haas S, de Kam M, et al. An overview of the CNS-pharmacodynamic profiles of nonselective and selective GABAA agonists. *Adv Pharmacol Sci.* 2012;2012:134523.
  69. Venkatakrishnan K, Culm KE, Ehrenberg BL, et al. Kinetics and dynamics of intravenous adinazolam, N-desmethyl adinazolam, and alprazolam in healthy volunteers. *J Clin Pharmacol.* 2005;45(5):529-537.
  70. Mould DR, DeFeo TM, Reece S, et al. Simultaneous modeling of the pharmacokinetics and pharmacodynamics of midazolam and diazepam. *Clin Pharmacol Ther.* 1995;58(1):35-43.
  71. van Steveninck AL, Schoemaker HC, Pieters MS, et al. A comparison of the sensitivities of adaptive

- tracking, eye movement analysis and visual analog lines to the effects of incremental doses of temazepam in healthy volunteers. *Clin Pharmacol Ther.* 1991;50(2):172-180.
72. Lommerse J, Plock N, Cheung SYA, Sachs JR. V<sup>2</sup> ACHER: Visualization of complex trial data in pharmacometric analyses with covariates. *CPT Pharmacometrics Syst Pharmacol.* 2021;10(9):1092-1106.
73. Singh RSP, Toussi SS, Hackman F, Chan PL, Rao R, Allen R et al. Innovative Randomized Phase I Study and Dosing Regimen Selection to Accelerate and Inform Pivotal COVID-19 Trial of Nirmatrelvir. *Clin Pharmacol Ther.* 2022;112(1):101-111.
74. Chan P, Yu J, Chinn L, Prohn M, Huisman J, Matzuka B et al. Population Pharmacokinetics, Efficacy Exposure-response Analysis, and Model-based Meta-analysis of Fenebrutinib in Subjects with Rheumatoid Arthritis. *Pharm Res.* 2020;37(2):25.
75. The European Agency for the Evaluation of Medicinal Products: Evaluation of Medicines for Human Use: Committee for Proprietary Medicinal Products (CPMP). Points to Consider on Applications with 1.Meta-analyses; 2.One Pivotal Study. [https://www.ema.europa.eu/en/documents/scientific-guideline/points-consider-application-1meta-analyses-2one-pivotal-study\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/points-consider-application-1meta-analyses-2one-pivotal-study_en.pdf) [アクセス日: 12 Mar 2026].

## データサイエンス部会 2025年度継続タスクフォース 4

### タスクフォース推進委員

MSD 株式会社	長谷川 千尋
ファイザーR&D 合同会社	松岡 伸篤

### タスクフォースメンバー

エーザイ株式会社	早戸 誠一
ブリistol・マイヤーズ スクイブ株式会社	田中 潤
ファイザーR&D 合同会社	吉松 宏樹
塩野義製薬株式会社	勝部 孝行
住友ファーマ株式会社	根本 大輔
サノフィ株式会社	渡瀬 広崇
大塚製薬株式会社	佐々木 智啓 リーダー

### 監修

担当副部会長 菅波 秀規	興和株式会社
--------------	--------

以上の資料作成にあたり、本資料の査読を実施頂いた査読担当者の諸氏に感謝いたします。