



医療技術評価における多変量メタアナリシスと代替エンドポイント

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年度継続タスクフォース7

医療経済評価に関連する統計的手法の整理と紹介

Ver 1

2023年4月

本報告書の中で Technical Support Document 20 (Decision Support Unit, National Institute for Health and Care Excellence) を引用／参考にした記載は、本タスクフォースの解釈であり、NICE には一切の責任はない。また TSD 20 の全てを紹介したわけではなく、一部を抜粋し紹介している。それらの紹介にあたり、タスクフォースによる解釈や追加の説明を加えている。これらは本タスクが独自にまとめたものであり、NICE から見解を得たものではない。

目次

1. はじめに.....	3
2. 概要.....	6
3. 多変量メタアナリシスへのベイズ流アプローチ.....	8
3.1 試験内の分散共分散行列.....	10
3.2 試験間の分散共分散行列.....	11
3.3 多変量メタアナリシスの利点.....	12
3.4 技術評価での2変量メタアナリシスの事例.....	15
4.2 変量メタアナリシスによる代替エンドポイントの評価.....	17
4.1 代替エンドポイント, それらの重要性と妥当性.....	17
4.1.1 代替エンドポイントの重要性.....	17
4.1.2 代替エンドポイントの妥当性.....	20
4.1.3 代替エンドポイント評価におけるメタアナリシス.....	21
4.1.4 代替エンドポイント検証のデータ要件.....	22
4.2 標準的な代替性のモデル (Daniels and Hughes).....	23
4.3 積・正規分布配合におけるBRMA.....	24
4.4 標準形式のBRMA.....	24
4.5 文献の他のモデルとの関係.....	24
4.6 検証と予測.....	26
4.6.1 クロスバリデーションの手順.....	26
4.6.2 新しい試験における代替エンドポイントの治療効果による, 真のエンドポイント に対する治療効果の予測.....	27
4.7 例: 再発寛解型多発性硬化症における代替性の妥当性評価.....	28
4.8 代替性の基準, 他の代替エンドポイントモデル及びさらなる研究の議論.....	28
4.9 その他の発展的なモデル.....	29
4.10 推奨事項.....	32
5. 多変量変量効果メタアナリシスのモデル.....	34
6. 議論と拡張.....	43
6.1 変量効果の正規性の仮定.....	43
6.2 二値変数, ネストされた, 又は相互に排他的なアウトカム.....	43
6.3 ベイズの方法と頻度論の方法, 及び関連ソフトウェア.....	44
6.4 その他の応用分野 (QOL 値など).....	44
7. まとめ.....	45
8. 引用文献.....	46
付録.....	55

1. はじめに

“宇沢弘文によれば、ゆたかな社会とは「すべての人々が、その先天的、後天的資質と能力とを十分に生かし、それぞれのもっている夢とアスピレーション (aspiration, 熱望, 抱負) が最大限に実現できるような仕事にたずさわり、その私的、社会的貢献に相応しい所得を得て、幸福で、安定的な家庭を営み、できるだけ多様な社会的接触をもち、文化的水準の高い一生をおくることができるような社会」である。そのためには、第一に「美しいゆたかな自然環境が安定的、持続的に維持されている」こと、第二に「快適で、清潔な生活を営むことができるような住居と生活的、文化的環境が用意されている」こと、第三に「すべての子どもたちが、それぞれのもっている多様な資質と能力をできるだけ伸ばし、発展させ、調和のとれた社会的人間として成長しうる学校教育制度が用意されている」こと、第四に「疾病、傷害にさいして、そのときどきにおける最高水準の医療サービスを受けることができる」こと、そして第五に「さまざまな希少資源が、以上の目的を達成するためにもっとも効率的、かつ公平に配分されるような経済的、社会的制度が整備されている」ことが満たされていなければならないと指摘する。 [1][2]”

日本を代表する経済学者のひとりである故・宇沢弘文が考える「ゆたかな社会」とは、自然環境、生活環境、文化的環境、及び学校教育の豊かさに加え、医療環境の豊かさでした。具体的には、最高水準の医療サービスの享受、その貴重な医療資源を効率的、公平に配分できる制度の整備でした。医療資源を効率的に配分することを目指したのが医療技術評価であり、その一部を構成するものが費用対効果評価です。費用対効果評価の制度の整備に関しては、既に多くの国で医薬品や医療機器の保険償還又は薬価調整等に用いられており、本邦では2019年より医薬品及び医療機器の価格調整を目的として、費用対効果評価制度が本格導入されました[3]。英国では1999年にNICE (National Institute for Health and Care Excellence) が設立され、医療技術等の保険償還の可否、及び新薬の薬価に関連した費用対効果評価及び技術評価を実施しており、制度のみならず、評価の事例や技術的側面についても参考になるものがあります[4]。NICEのDSU (Decision Support Unit) は、様々な技術的文書 (Technical Support Document, TSD) を作成しており、現在までに21のTSDが発行されています[5]。本タスクフォースでは、2018年より、NICEのTSDで扱われるトピックスを整理して、国内で費用対効果評価を実践する際に有用と思われる情報を提供する活動を行ってきました。

費用対効果を検討するにあたっては、評価対象技術の比較対照技術に対する追加的な有用性をシステマティックレビューによって評価する必要があります。しかしなが

ら、新しい医療技術（新薬など）の場合、費用対効果評価での対照技術（既存の治療薬など）を直接比較したエビデンス（ランダム化比較試験の結果など）がないこともあり、そのような場合には、間接比較が用いられることがあります。その際、各試験での治療効果（プラセボ群と比較した治療効果の推定値など）を単に引き算するなどといった単純な間接比較は推奨されません。三治療以上の治療技術を含めて間接比較を行う場合には、評価対象技術やその比較対照技術が含まれていない試験も包括的に活用できるネットワークメタアナリシス（NMA）と呼ばれる手法が検討されます。NMAは、一定の前提条件の下で、包括的にエビデンスを活用し試験間に存在する交絡因子を調整できることから、単純な間接比較に比してバイアスの減少及び推定精度の向上が期待できます。本タスクフォースでは、2020年にTSDのEvidence synthesisシリーズ（TSD 1～7）をもとに、NMAを中心としたエビデンス統合の方法論及び留意事項を要約し、各TSDの付録のWinBUGSの使い方と解析事例も交えて解説した報告書「NICE TSD Evidence Synthesis シリーズの紹介：ネットワークメタアナリシスを中心として」を作成しました[6]。

本報告書では、TSDのEvidence synthesisシリーズの続報ともいうべき、多変量メタアナリシス（multivariate meta-analysis）と代替エンドポイント¹に関連した方法論を扱うTSD 20: Multivariate meta-analysis of summary data for combining treatment effects on correlated outcomes and evaluating surrogate endpoints [7]を紹介します。TSD 20では、多変量メタアナリシスの主な方法と、多変量メタアナリシスを多数の治療を比較する多変量ネットワークメタアナリシス（multivariate network meta-analysis）へ拡張する手法を扱います。特に全てのアウトカムを報告していない試験が存在する場合でも、利用可能な全ての注目するアウトカムをモデルに含めることで、アウトカム間で情報を借用することが可能となります。また、多変量メタアナリシスの代替エンドポイントの評価への応用も扱います。TSD 20は、2022年12月に第2版に改訂されました。改訂内容は、三変量変量効果メタアナリシスで試験間分散行列をunstructuredとした場合の正規分布の積による定式化（product normal formulation, PNF）による実装（5章参照）に関する数式の修正です。

新しい医療技術の薬事承認に代替エンドポイント（例えば、全生存期間よりも早期に評価可能な指標としての無増悪生存期間）が用いられることも増えています[8][9][10][11][12][13]。薬事承認段階で真のエンドポイントに関するデータが限定的にしか得られていない場合、もしくは全く得られていない場合でも、長期の有用性や経済性を評価することが求められ医療技術評価（health technology assessment, HTA）の

¹ 本書では誤解が生じない限り、エンドポイントとアウトカムを交換可能な用語として使用しています。

プロセスでは、代替エンドポイントと真のエンドポイントとの関係をモデリングし、経済モデルに含めることが推奨されます[14]。その際に、真のエンドポイントに対する治療効果をよりよく推定することができれば、分析結果の不確実性を減少させることができます。多変量メタアナリシスは、複数アウトカムに対する治療効果を推定し、相関や不確実性を考慮した分析を行うための枠組みを提供します。

本報告書では TSD 20 で扱われた、ベイズ流の枠組みによる多変量メタアナリシスの手法を紹介します。TSD 20 で紹介しているベイズ流の多変量メタアナリシスは、代替アウトカムのエビデンスは得ているが、真のエンドポイントのエビデンスが得られていない段階で新規の薬剤の医療技術評価を行う場合や、複数のアウトカムを用いて経済モデルを構築する場合により多くのエビデンスを活用するといった際に有用なアプローチです。後者の場合は、強度の借用（後述）により治療効果の推定精度の向上も期待できます。

最初に、医療技術評価と多変量メタアナリシスの概要を説明し、ベイズ流多変量メタアナリシスの導入を行います（数理的なモデルの詳細は 5 章にまとめています）。ベイズ流多変量メタアナリシスに特徴的な試験内及び試験間の分散共分散行列の扱いと多変量メタアナリシスの利点を紹介し、英国 NICE での技術評価（technology appraisal, TA）での事例を紹介します。次に、代替エンドポイントの評価への多変量メタアナリシスの応用を扱い、課題と発展的なモデルも紹介したうえで、代替エンドポイント評価に対する NICE の推奨事項と NICE での評価事例を紹介します。TSD 20 では、関節リウマチと多発性硬化症の事例を用いて、各手法の WinBUGS による実装が紹介されていますが、本報告書では、数値例は割愛しておりますので、ご興味のある方は原著をご参照ください。

マルコフ連鎖モンテカルロ（Markov chain Monte Carlo, MCMC）を用いるベイズ流メタアナリシスの WinBUGS による実装は、TSD 2~5 で導入されており[15][16][17][18]、製薬協の報告書にも紹介がありますので、併せてご参照ください[6]。ベイズ流メタアナリシスは、WinBUGS に限らず SAS の MCMC プロシジャや Stan でも実行可能です。参考までに、代替エンドポイントを評価する Daniels and Hughes モデルを実装した WinBUGS コード（プログラム B.1）を SAS および Stan に移植したコードを付録に示しています。

本報告書は医療経済分析の基本的な知識を前提にしています。費用対効果評価制度の背景や費用効果分析の基本的な考え方等については、保健医療経済評価研究センター（CORE 2 Health）のサイト、中医協ガイドライン第 3 版、日本製薬工業協会の過去の報告書、成書等を必要に応じて参照してください[19][14][20][21][22][23][24]。

2. 概要

HTAにおいて、エビデンスに基づく意思決定を行うためには、評価対象技術の比較対照技術に対する追加的な有用性を系統的文献レビューにより収集し、その結果をメタアナリシス等の手法を用いて適切に統合する必要があります。例えば、新規医療技術の費用対効果評価は、新治療の有効性（effectiveness）に関するメタアナリシスの結果に大きく依存します。意思決定の基盤となるエビデンスは、通常、ランダム化比較試験（randomized controlled study, RCT）の系統的文献レビューから得られます。各試験で報告されている臨床アウトカムには多くの異質性（heterogeneity）があります。原因は、有効性の評価尺度や評価時点やアウトカムの定義の違い、選択的な報告、主要なアウトカム間の結果の不整合などです。加えて、治療効果の評価に長い追跡期間が必要なエンドポイントの場合、代わりに代替エンドポイントの結果が報告されているRCTがあります。こうした問題は、経済モデルによる推定に利用可能なエビデンスの範囲を制約するため、HTAの意思決定における大きな課題です。

ベイズ流のアプローチを用いた単一アウトカムのメタアナリシス及びネットワークメタアナリシスは、既報の「NICE TSD Evidence Synthesis シリーズの紹介：ネットワークメタアナリシスを中心として」で紹介しています[6]。メタアナリシスは頻度論のアプローチも存在しますが、ベイズ流のアプローチには次の利点があるとされています。パラメータの事後分布の確率的な解釈が可能であり、分析に含めたある治療が最も優れる確率を計算することができる点や情報源の不確実性を反映する点、分析ツールが充実している点が挙げられます[25]。また、ベイズ流アプローチは、複雑なデータ構造をモデル化するための柔軟な枠組みとなり得ます。ネットワークメタアナリシス（network meta-analysis, NMA）は、複数の治療オプションを同時に比較することができます。単変量のメタアナリシスを多変量に拡張した多変量メタアナリシスは、相関のある複数のアウトカムに対する治療効果を同時にモデル化することができます。NMAはすでにHTAでランダム化比較試験から得られたエビデンスを合成するための標準ツールになっています。

HTAの追加的有効性の評価を行う際に、複数のアウトカムの報告がある場合でも、メタアナリシスによるエビデンスの統合がアウトカムごとに行われることがしばしばあります。つまり、アウトカムごとに標準的な単変量メタアナリシスを個別に行いますが、このとき注目するアウトカムがない試験は、そのアウトカムのメタアナリシスから除外されてしまいます。しかし、試験の除外は望ましいことではありません。特に試験参加者が検討の対象集団や臨床的な状態を代表する場合には、そうした試験を除外すべきではありません。臨床試験には多くの費用と時間が必要であり、患者の参加

も貴重なものです。その試験に含まれる他のアウトカムが注目するアウトカムに対して（間接的な）情報を持つ場合に、単純な試験の除外はエビデンスの破棄にもなります。また、臨床アウトカムは互いに相関するものも多くあります [26]。例えば、高血圧症患者の収縮期血圧と拡張期血圧、片頭痛の痛みと悪心のレベル、がん患者の無病生存期間（disease-free survival, DFS）と全生存期間（overall survival, OS）には相関があります。個人レベルでのエンドポイント間の相関は、母集団あるいは試験レベルでの治療効果間の相関をもたらします。例えば、降圧治療のランダム化試験では、収縮期血圧と拡張期血圧に対する治療効果の推定値は高度に相関している可能性が高いでしょう。同様に、がんの試験では、腫瘍の縮小など予後バイオマーカーに対する効果の推定値は、無病生存率や全生存率と相関があるでしょう。例えば、複数の時点（経時測定データ） [27]、複数のアウトカムに関連する共通の遺伝的要因 [28]、複数のアウトカムについて共通の共変量を用いて得られた推定値 [29]、精度やパフォーマンスに関する複数の測定値（例えば、診断テストや予測モデル） [30]、同じ構成要素から成る複数の測定値（例えば、異なる痛みのスケールからのスコアや異なる測定方法で得られたバイオマーカー値 [31] [19]）などです。ここでは、これらをまとめて相関する複数のアウトカムと呼ぶことにします。

多変量メタアナリシスは、試験内及び試験間相関を利用して複数のアウトカムを同時に分析することで、より多くの試験を推定に貢献させることができます。単変量メタアナリシスとの比較で推定効率が向上するのみならず、バイアス（例えば、アウトカムの選択的な報告によるバイアス）を減少させる可能性があります [32]。各アウトカムの推定値には、直接的なエビデンスだけでなく、相関する他のアウトカムのエビデンスも寄与します。多変量メタアナリシスは、より多くの試験をエビデンスの基盤とするための自然な枠組みを提供します。特にメタアナリシスに含める試験数が少ない場合や注目するアウトカムのデータが欠測している場合に、プールした治療効果の推定精度を向上できる可能性があります。また、最近の方法論の著しい進展に伴い、徐々に用いられるようになってきました [33] [34] [35] [36] [37] [38]。多変量メタアナリシスの利用の増加には、代替エンドポイントの評価を含むエビデンス統合において、代替エンドポイントの治療効果と、最終的な臨床アウトカムの治療効果を同時にモデル化することで、臨床的有効性の推定値を早期に取得できるという利点があることがその理由のひとつです。

このあと、多変量メタアナリシスを実装するために TSD 20 で扱うベイズ流の枠組みを紹介し、この方法論が有用なシナリオを説明します。

3. 多変量メタアナリシスへのベイズ流アプローチ

多変量アプローチを使用して複数のアウトカムのデータを統合することの利点のひとつは、アウトカム全体でエビデンスの強度を借用することで (borrowing of strength, BoS), 得られる有効性の推定値の不確実性を減少させる可能性があることです。BoSは、分析モデルの仮定の下で、関心のあるパラメータ以外の他のデータを、パラメータ間の相関を通じて、間接的に推定に利用することを指します[39]。ベイズ流の枠組みでは、幅広いエビデンスソースを分析に含めることができます[40][41][42]。TSD 20で具体的に扱っていないケースも含めると、例えば、多変量メタアナリシスには、追加の情報（観察研究や臨床試験、系統的レビューなどの外部データ）や専門家の意見を事前分布の形で取り入れることができます。外部データの使用により、推定値に関する不確実性が減少する可能性があります。ベイズ流アプローチでは、モデルパラメータについて、例えば、対照治療に対する新治療のOSやPFSのハザード比は1未満であるというような情報を、直接確率ステートメント (direct probability statement) の形で記述することもできます。

ここでは、ベイズ流の枠組みにより多変量メタアナリシスを行います。具体的には、式(1), (2)で表される階層ベイズモデルによるモデリングを行います。式(1)は、試験内の階層で式(2)は試験間の階層（ハイパーパラメータ）です。それぞれ、正規分布によりモデル化します（オッズ比、ハザード比などは変数変換により正規分布に近似します）。

$$Y_i \sim N(\delta_i, \Sigma_i) \quad (1)$$

$$\delta_i \sim N(d, T) \quad (2)$$

Y_i は、試験*i*のアウトカムベクトル、このアウトカムベクトルは特に断らない限り、比較対照との差や比の治療効果を表すことに注意してください。試験*i*には複数個のアウトカムの結果があります。 Σ_i は試験*i*のアウトカム間の分散共分散行列です。 Y_i が治療効果の推定値の場合、臨床試験結果の論文等で分散の推定に関する情報である信頼区間や標準誤差はほぼ報告されるのに対し、複数のエンドポイント間の相関に関する情報は報告されないことがほとんどです（試験の総括報告書にも存在しないことが多いでしょう）。試験内の相関は通常、推定の興味の対象ではありませんが（局外パラメータ）、階層モデルを用いた推定のためになんらか値（又は分布）が必要です。その推定値を得ることは、階層モデルの課題のひとつであり、その方法のいくつかを後に紹介します。

δ_i は、試験*i*の真の治療効果を表す確率変数ベクトルで、正規分布に従うと仮定します。正規分布の平均値 d は、対照治療に対する治療効果を表します。メタアナリシスの第一の目的は、 d を推定することです。 δ_i の相関 T は、あるエンドポイントが他のエンドポイントの結果に関する情報をもつ程度を表すという点で、代替エンドポイントの評価において重要です。また、この相関が大きければ一方のエンドポイントに対する治療効果を他方のエンドポイントの治療効果の推定に寄与する度合いが大きくなります。これは二変量（多変量）のネットワークメタアナリシスにおいて特に重要です。

この階層モデルでは、試験レベルの効果は交換可能（exchangeable）と仮定します。つまり、 δ_i は試験によらず同様（similar）、すなわち試験間で交換可能な δ_i は共通の正規分布 $N(d, T)$ に従うと仮定します[15]。

さらに、この分散行列 T の構成要素である分散及び相関係数にも交換可能性を仮定することで（各エンドポイントの分散、及びエンドポイント間の相関が試験間で同じ。 T が均一（homogeneity））、推定に必要なパラメータ数を削減できると共に、更に情報を借用することが可能となります。これらの交換可能性が疑われる場合（例えば、作用機序の異なる複数の薬剤クラスを含むネットワークメタアナリシス）、交換可能性を仮定しないモデルを用いることもできます[43]。

各試験の治療効果（ δ_{i1}, δ_{i2} ）の推定値は、他の試験の情報を借用したベイズ流の縮小推定（"shrunk" estimates）になります[44]。

多変量メタアナリシスを 3 以上の治療を含めるネットワークメタアナリシスへ拡張する研究も進んでいます[45][46][47][48][49][50]。単変量のネットワークメタアナリシスの方法論は主に TSD 2 で扱っており、製薬協の報告書でも解説されています[15][6]。単変量ネットワークメタアナリシスの場合と同様に、一致性（consistency）を仮定します。一致性とは、治療 A と C の比較と治療 B と C の比較から矛盾なく治療 A と B の比較が得られる、すなわち移行性（transitivity）が成立するという仮定です。具体期には、治療 k, b を比較した試験 i のアウトカム j を $Y_{(bk)ij}$ とし、試験レベルのモデルを考えます。

$$\begin{pmatrix} Y_{(bk)i1} \\ Y_{(bk)i2} \end{pmatrix} \sim N \left(\begin{bmatrix} \delta_{(bk)i1} \\ \delta_{(bk)i2} \end{bmatrix}, \begin{bmatrix} \sigma_{(bk)i1}^2 & \sigma_{(bk)i1}\sigma_{(bk)i2}\rho_{(bk)wi} \\ \sigma_{(bk)i1}\sigma_{(bk)i2}\rho_{(bk)wi} & \sigma_{(bk)i2}^2 \end{bmatrix} \right)$$

ここで、治療効果 δ_i について、次式に示す移行性を仮定します。

$$\begin{pmatrix} \delta_{(kl)i1} \\ \delta_{(kl)i2} \end{pmatrix} = \begin{pmatrix} \delta_{(bl)i1} - \delta_{(bk)i1} \\ \delta_{(bl)i2} - \delta_{(bk)i2} \end{pmatrix}$$

このとき、母数レベルのモデルは次のようになり、

$$\begin{pmatrix} \delta_{(bk)i1} \\ \delta_{(bk)i2} \end{pmatrix} \sim N \left(\begin{bmatrix} d_{(bk)1} \\ d_{(bk)2} \end{bmatrix}, \begin{bmatrix} \tau_{(bk)1}^2 & \tau_{(bk)1}\tau_{(bk)2}\rho_{(bk)} \\ \tau_{(bk)1}\tau_{(bk)2}\rho_{(bk)} & \tau_{(bk)2}^2 \end{bmatrix} \right)$$

δ_i についての移行性から、一次の一致性（first order consistency equation）が導かれます。

$$\begin{pmatrix} d_{(kl)i1} \\ d_{(kl)i2} \end{pmatrix} = \begin{pmatrix} d_{(bl)1} - d_{(bk)1} \\ d_{(bl)2} - d_{(bk)2} \end{pmatrix}$$

TSD 20 では、2 つの治療の 2 変量メタアナリシスの直接的な拡張である治療の対比（群間の差や比）をインプットとした 2 変量のネットワークメタアナリシスから始めて、対比ではなく群ごとの治療効果の推定値をインプットとした 2 変量のネットワークメタアナリシスを説明し、さらにこれを多変量のネットワークメタアナリシスへ拡張するという順に数値例を交えて解説されています（本報告書では、モデルの数理的な側面は 5 章にまとめています）。

3.1 試験内の分散共分散行列

複数のアウトカムに対する治療効果の試験内相関は、同じ患者で複数のアウトカムが測定されるために発生します。多変量メタアナリシスでは、試験内相関を適切に考慮する必要があります。試験内相関を無視した場合、統合された効果の推定値の不確実性が大きくなり、バイアスを含んだものとなる可能性があります [51]。

最も単純なモデルでは、試験内共分散行列 Σ_i の各成分は既知として扱います。試験内分散の推定値は、通常、公表された試験報告から各アウトカムに対する治療効果の標準誤差の二乗として得られますが、2 つのアウトカムの間の試験内相関は報告されないことがほとんどなので、推定値を得ることが困難です。患者レベルのデータ

（individual patient data, IPD）が利用可能であれば、ブートストラップ法[52]や誤差に相関のある 2 つのアウトカムに対する回帰モデルによって[31]、相関を推定することができます。連続的なアウトカムの場合、後者のアプローチが自然です。他方、二値変数やイベント時間変数など、別の種類のアウトカムの場合、回帰によるアプローチは複雑なものとなるため、ブートストラップ法を用いる方がよいでしょう。ブートストラップ法では、元データのアウトカムのセット（例えば、二値変数と生存時間変数のセット）からの復元再抽出により、元データセットと同じ大きさのデータセットを複数個（例えば 1,000 個）生成します。相関構造を維持するために、複数のアウトカムはセットで（試験参加者の単位で）抽出します。各データセットで、メタアナリシスで用いるのと同じ要約データのセット（例えば、二値変数の対数オッズ比と生存時間変数の対数ハザード比のセット）をそれぞれ算出します。こうして、ブートストラップ標本から対数オッズ比と対数ハザード比の 1,000 個のペアを作り、その相関を計

算します。これが、2つのアウトカムに対する治療効果間の試験内相関の推定値となります。

多くの場合、IPDが利用できるのはごく一部の試験に限られるか、あるいはIPDは全く利用できません。IPDが、例えば1つの試験でしか利用できない場合、治療効果間の相関は全ての試験で同じであると仮定することもできます。ベイズ流の枠組みでは、試験内相関の推定値に関する不確実性に対応するため、試験内相関の事前分布を与えることができます。例えば、一部の試験のみで利用可能な場合に、試験内相関を欠測値とする代わりに、事前分布を与えることができます。試験内相関の事前分布は、例えば、IPD（メタアナリシスに含まれる試験あるいは同様の集団を対象とした外部の試験）の二重ブートストラップ法により構成できます。二重ブートストラップ法では、最初の実データからのブートストラップ標本を抽出し、二段階目は、一段階目で得たブートストラップ標本からさらにブートストラップ抽出を行います。単純なブートストラップ法はひとつの相関を得ますが、二重ブートストラップ法では相関の分布を得ます。相関の分布により不確実性を表します。この手法は、Bujkiewiczらによって実装されました[37]。

IPDが得られない場合に検討可能なアプローチもいくつかあります。単純な方法は、試験内相関の値の範囲（-1～1）を感度分析で検討することです。試験内相関を推定するアプローチに、Wei and Higginsによって提案された方法があります。それは、2つのアウトカムに対する治療効果間の試験内相関を、報告されていることが多いであろう測定値間の相関の関数で表す方法です。例えば、治療効果の2つの対数オッズ比の相関を、2つの二項変数間の相関から求めます[53]。他にも、Rileyら[54]によって提案された、試験内相関が不明な場合の2変量メタアナリシスの手法も利用可能です。その手法は、試験内モデルと試験間モデルの両方の共分散をひとつの項に結合することで、試験内相関の指定を不要にします。Rileyらの手法の限界は、結合された項は、試験内相関としても試験間相関としても解釈できないことです。そのため、例えば代替エンドポイント評価での利用に制約が生じます。

3.2 試験間の分散共分散行列

三つ以上のアウトカムの多変量メタアナリシスの複雑さは、試験間の分散共分散行列の事前分布にあります。分散共分散行列が半正定値（positive semi-definite）行列となることを保証するような方法で、事前分布を与える必要があります。2変量の場合は、分散共分散行列の各要素が取り得る値（分散は正、相関は-1と1の間）に制限するように事前分布を与えれば十分でした。多変量の場合、分散共分散行列の要素に独

立した事前分布を与えると、MCMC シミュレーションでサンプリングされた値による行列が、半正定値とならない場合があります。試験間の分散共分散行列に適切な事前分布を与える方法はいくつかあります。例えば、TSD 20 の付録 C.1 で用いているコレスキー分解 (Cholesky decomposition) や球面分解 (spherical decomposition) はそのひとつです。分散共分散行列を半正定値とするために、分散共分散行列の事前分布に共役事前分布である逆ウィシャート分布 (inverse Wishart distribution) を採用することも考えられますが、無情報な事前分布を構成することが難しいことと、事後分布への影響から TSD 20 では推奨されていません[6][36]。

試験内相関とは違って、試験間相関は試験レベルの要約データから推定することができます。ベイズ流の枠組みを用いる場合のひとつの方法は、試験間相関には無情報事前分布を与えることです。事前分布の詳細は TSD 20 の 2.1.2 節をご参照ください。もうひとつの方法は、ベイズ流アプローチの利点を生かし、外部情報を用いて情報をもつ事前分布を構成することです。後者の詳細は TSD 20 の 2.4 節及び Bujkiewicz らの報告を参照ください[37][56]。

一部の試験では一部のエンドポイントの結果だけが利用可能な場合があります、当該試験で利用可能でないエンドポイントの治療効果、対応する標準誤差が欠測値となります。その場合、missing at random (MAR) の仮定のもと、事後分布の予測値より、他の試験からの情報を借りる形で、欠測部分を補完できます。この場合、母集団の治療効果の分散 (標準偏差の 2 乗) $v_{ji}(= \sigma_{ji}^2 N_i)$ (すなわち、 i 試験の j エンドポイントの治療効果の分散) を交換可能とし、 v_{ji} は同一分布から得られるものとし、得られた v_{ji} にもとづき、治療効果の標準誤差を $\sigma_{ji}^2 = v_{ji}/N_i$ (N_i : 総症例数) の平方根として推定します。試験内の治療効果に対する分散 σ_{ji}^2 の交換可能性は、 σ_{ji}^2 が試験のサイズ N_i に依存するため、成立しないことに注意が必要です。一方、A、B 群で症例数が異なる場合は、 v_{ji} の意味合いを変え、母集団の各群共通の分散として生成され、治療効果の標準誤差は $\sigma_{ji}^2 = v_{ji}(1/N_{Ai} + 1/N_{Bi})$ の平方根として求めます。 v_{ji} の事前分布の詳細は、TSD 20 の 2.1.3 節をご参照ください。代替法として、欠測となっているアウトカムに対し任意の治療効果を入力し、非常に大きい分散と、共分散 0 を入力して、MCMC で推定する方法もあります。

3.3 多変量メタアナリシスの利点

Riley らの議論によると、多変量メタアナリシスの利点には報告バイアスの減少も挙げられますが、最も大きなものは治療効果の推定の精度の向上に関連するものです

[26]. 精度の向上は、効率の尺度[57]と強度の借用 (BoS) [39]によって定量化することができます。Copas らは、同様の試験間の異質性がある場合の多変量 (あるいはネットワーク) メタアナリシスと比較して、標準的な単変量のメタアナリシス (NMA の場合は直接比較に基づくエビデンスのみ) が利用可能な試験の全てのエビデンスの $100 \times (1-E)\%$ を無駄にしているという評価指標を提案しました[57]. 効率の指標 E は次のように定義されます。

$$E = \frac{\text{直接比較及び関連するエビデンスに基づく推定値の分散}}{\text{直接比較エビデンスのみに基づく推定値の分散}}$$

ここで、関連するエビデンスは、間接比較に基づくエビデンスと相関に基づくエビデンスのいずれか又は両方を指します。分散は、メタアナリシスで推定するパラメータの本来のスケールの分散です (よくあるのは、リスク比の対数、オッズ比の対数、ハザード比の対数、平均値の差)。例えば、 $E=0.9$ の場合、標準的なメタアナリシスは、全ての試験の全ての患者 (又はイベント) の 10% を無駄にしていることに相当します。また、 n を、直接比較のエビデンスとして利用可能な試験 (標準的なメタアナリシスに貢献すると考えられる試験) の数と定義します。そして、間接比較又は相関に基づくエビデンスを利用することによる情報の増加を、 n 個の試験が似たサイズであるとして、 $n \times ((1-E)/E)$ で算出します。例えば、あるアウトカムについての 9 試験の標準的なメタアナリシスを行い $E=0.9$ であった場合、多変量メタアナリシスによる情報の増加は、 $9 \times (1-0.9) \div 0.9 = 1$ 試験分となります。すなわち、多変量メタアナリシスによって、およそ 1 試験に費やした時間、労力、費用が得られたと考えることができます。

強度の借用の統計量 $BoS = 100 \times (1-E)\%$ も Jackson らによって提案されており[39], エビデンスの借用による推定値の分散の減少をパーセントで表すものです。また、 BoS は、相関や間接比較に基づくエビデンスが推定に寄与する割合と解釈することもできます[58]。例えば、ネットワークメタアナリシスで、 0% の BoS は要約された結果が直接比較のエビデンスのみに基づくことを、 100% の BoS は完全に相関や間接比較に基づくエビデンスに依存していることを意味します。Riley らは、ネットワークメタアナリシスや複数アウトカムのメタアナリシスといった多変量メタアナリシスで、各試験の寄与を算出する手法を提案しています[59]。

Riley らに倣うと、複数のアウトカムの多変量メタアナリシスの潜在的な重要性は、情報の借用と精度の向上が大きいとき最大になります[26]。次のような場合です。

- 注目するアウトカムの直接エビデンスがない試験の割合が大きい
- 注目するアウトカムが報告されていない試験で、他のアウトカムの結果が利用可能
- アウトカム間（試験内もしくは試験間）の相関が高い（例えば、 > 0.5 又は < -0.5 ）

Reily らは、通常、複数治療のネットワークメタアナリシスで、BoS が最大になると主張しました[26]。つまり、経験に基づくと、通常、複数治療の多変量メタアナリシスでは、一致性の仮定の下で間接比較による情報の増加は、アウトカム間の相関による情報量の増加よりも大きいのです。複数アウトカムの多変量メタアナリシスは、相関の高いアウトカムに適しています。それ以外の場合、通常、BoS は小さくなります。BoS が大きくなるアウトカムの組合せを探すようなデータの「どぶさらい」を避けるために、相関の高いアウトカムは、臨床的な判断や統計的な知見などから、メタアナリシスを行う前に特定しておくべきです。複数アウトカムの多変量メタアナリシスは、試験レベルでアウトカムが欠測している場合にも適しています。次の事例が示すように、あるアウトカムに対する BoS は、そのアウトカムが欠測している試験の割合がおよその上限となります。例えば、心血管疾患の発現率へのフィブリノーゲンが 1 単位増加することの予後に対する効果を統合するメタアナリシスの例では、アウトカムが欠測した試験が 31 試験中 17 試験、55%であり、欠測が完全に調整された場合、BoS は最大 55%になるため、多変量メタアナリシスが選択肢になりました。実際に多変量ネットワークメタアナリシスを行ったところ BoS は 53%であり、理論上の最大値に非常に近い値でした[26]。これは、部分的に調整された効果と完全に調整された効果がほぼ完全に相関するためです。他方、欠測がなく完全にデータがある場合や欠測の割合が小さい場合、BoS は重要ではないでしょう。欠測が存在し BoS が大きくなりそうな場合、多変量メタアナリシスが有力な選択肢となります。完全なデータが存在する場合（全てのアウトカムが全ての試験で利用可能な場合）、BoS は通常、小さな値になります（例えば、 $< 10\%$ ）。また、多変量メタアナリシスでも対象とするアウトカムが一つもない試験は除外されます。すなわち、多変量メタアナリシスは、報告バイアス（reporting bias）について、選択的なアウトカムの報告によるバイアスを減少させることはできますが、試験が全く報告されないことに起因するバイアスには対処できません。

精度の向上に加えて、多変量メタアナリシスで使用される追加情報は、単変量メタアナリシスと異なる統合効果の推定値を与える可能性があります。これは、一部のアウトカムで選択的にデータが欠測している状況で発生する可能性が最も高くなります。Kirkham らは、一部のアウトカムは選択的に報告されていない（例えば、p 値に依存して）が、他の相関するアウトカムはおおむね利用可能という状況で、多変量メタアナ

リシスによって、アウトカムに関する報告バイアスを減少できると結論しました[32].

相関する治療効果の正式な比較が重要な場合（例えば、収縮期血圧と拡張期血圧に対する治療効果の違いを推定する）、欠測データの量に関係なく、多変量の枠組みで行うべきです[36]. 治療効果間の相関を考慮し、不確かな信頼区間と p 値を回避することができます。同様に、全ての試験が全ての治療を含む場合であっても、各治療の効果を推定し、比較するために単一の分析の枠組みを必要とする場合には、複数治療のネットワークメタアナリシスが推奨されます。

多変量（及びネットワーク）メタアナリシスの利点は、アウトカムの欠測メカニズムが、MAR であるという仮定に依存しています[60]. すなわち、いくつかの試験で観測された関係は、他の観測されなかった試験に移転可能（transferable）と仮定することになります。例えば、複数アウトカムの多変量メタアナリシスで、観測されたアウトカムの組に対する効果（試験内、試験間）についての線形な関連（相関）が、一方のアウトカムのみ利用可能な試験へ移転可能であると仮定します。これは、代替エンドポイントを正当化する際にも使われますが[61][55]、批判や論争の的になることもあります[62]. アウトカムの選択的な報告や選択的な解析については、missing NOT at random を仮定する方がより適切と思われまます[63][64]. MAR は、他の観測された変数で条件付けたとき、欠測値の分布が既知となることを仮定します。しかしながら、アウトカムが選択的に報告されている場合、MAR の仮定を成立させるために必要な全ての因子が得られることはないでしょう。それでも、多変量メタアナリシスを用いる際に、欠測値と関連する因子（相関のあるアウトカム）のひとつで条件付けるだけだとしても、単変量メタアナリシスよりも（統合した結果の統計的性質という意味で）良い結果が得られると TSD の著者は考えています。こうした状況において、多変量のアプローチは、完全ではないとしても、アウトカムが選択的に報告されることによるバイアスを減少させるのに役立ちます[32].

3.4 技術評価での 2 変量メタアナリシスの事例

NICE が実施した技術評価（technology appraisal, TA）での 2 変量メタアナリシスについての事例として、TSD 20 で紹介されている TA139, TA383 を紹介します。

閉塞性睡眠時無呼吸低呼吸症候群の治療のための持続的気道陽圧装置の技術評価では 2 変量メタアナリシスが使用されました[69]. この TA(TA139)では、臨床的有効性は複数のエンドポイント、エプワース眠気尺度（Epworth Sleepiness Scale, ESS）及び携帯型

血圧モニタリング（ambulatory blood pressure monitoring, ABPM）で評価されました。系統的文献レビューで特定された試験の一部のみが ABPM を報告していましたが、2変量メタアナリシスで2つのエンドポイントを一緒にモデル化することで、全ての試験をメタアナリシスに含めることができました。2変量メタアナリシスによる感度分析の結果は、従来の単変量メタアナリシスを使用して得られた結果と劇的に違うものではありませんでした。それでも、2変量メタアナリシスによる分析は報告バイアスの潜在的影響についての感度分析となるので有益です。つまり、報告された相関するアウトカムから情報を借用することで、未報告のアウトカムの影響が軽減され、また、結論が頑健（robust）かどうかを確認することができます。

2変量メタアナリシスのアプローチは、強直性脊椎炎（ankylosing spondylitis）及びX線基準を満たさない体軸性脊椎関節炎（non-radiographic axial spondyloarthritis）を対象とした腫瘍壊死因子 α 阻害剤（tumour necrosis factor- α inhibitors）のNICEの技術評価でも使用されました[66]。このTA(TA383)では、評価グループがエビデンス統合の一般的な枠組みを用いて、バス強直性脊椎炎疾患活動指数（Bath Ankylosing Spondylitis Disease Activity Index）によって測定される疾患活動性の変化を統合しました。その際に、バス強直性脊椎炎機能指数（Bath Ankylosing Spondylitis Functional Index）によって測定された機能に関する情報も統合しました。活動指数と機能指数のデータは2変量メタアナリシスの枠組みによって統合され、得られた治療効果の推定値とその不確実性が、経済モデルの主要な入力パラメータとなりました。

これらは医療機器や複数薬剤の公的分析の事例です。技術評価のみからでは分析の詳細までは読み取れないため、必要に応じて関連資料をご参照ください。

4.2 変量メタアナリシスによる代替エンドポイントの評価

本節では、2変量メタアナリシスを代替エンドポイントにより最終的な臨床アウトカム（以降、真のエンドポイント）を予測する問題に適用します。TSD 20ではベイズ流メタアナリシスによる予測を推奨していますが、過去の製薬協 医薬品評価委員会 統計・DM 部会資料にて、代替性の考え方の詳細（Prentice の基準、Freddman による Proportion of the Treatment Effect（PTE 法）、Buyse and Molenberghs らの評価指標の詳細、及び個別データ（IPD）メタアナリシスが紹介されています（現在非公開で、将来再公開の予定ですが、詳細は未定です）。また、Prentice の基準、Freddman の PTE、因果推論的アプローチ、及びメタアナリシスアプローチの各手法の特徴、事例、及びそれらの課題をまとめた総説[67]も、本書と併せてご参照頂ければと思います。

4.1 代替エンドポイント, それらの重要性と妥当性

4.1.1 代替エンドポイントの重要性

2変量及び多変量メタアナリシスの最も有用な応用のひとつに、代替エンドポイントの評価があります[68][52][69][70][71][55][72]。代替エンドポイントは、HTA 及び規制上の意思決定において重要性が増しています[73][74]。2変量メタアナリシスを利用し、代替エンドポイントの治療効果と、最終的な臨床アウトカムの治療効果を同時にモデル化することで、臨床的有効性の推定値を早期に取得できます。最終的な臨床アウトカムの結果が得られていない場合でも、代替エンドポイントの結果に基づいて新しい治療法に薬事承認が与えられることがあります。そうした未測定（又は他の理由で報告されていない）エンドポイントの治療効果を、代替エンドポイント（又は他の高度に相関するエンドポイント）の治療効果から予測し、意思決定モデルに用いることができます[75]。そうでなければ、有用性と外的妥当性を妥協した単純化したモデルを用いることになるかもしれません（補足：単純化しすぎたモデルでは、2変量メタアナリシスによる予測の有用性、及び外的妥当性が十分には享受できない可能性があるという主張と考えられます）。

欧州医薬品庁（EMA）や米国食品医薬品局（FDA）などの規制当局は、例えば代替エンドポイントで測定された治療効果に基づき、条件付き薬事承認を与えるなど、柔軟な薬事承認制度を導入しています[76][77]。最終的な臨床アウトカムでのデータが集積した時点で、薬剤は再評価されます。例えば、2008年1月から2012年12月までに、FDAは54の抗がん剤のうち36（67%）を代替エンドポイントに基づいて承認し、そ

のうち 19 は反応率 (response rate) , 17 は無増悪生存期間 (PFS) 又は無病生存期間 (disease-free survival) に基づく承認でした[78]. 条件付き薬事承認は, 希少疾患に対する薬剤 (稀な, 生命を脅かす, 又は慢性的に衰弱させる状態に用いられる薬剤) などに与えられる場合が多く, この場合, 臨床試験の患者数が少なく, 治療効果の早期の結果には不確実性が伴います. 例えば, 進行軟部肉腫の患者に使用されているオララツマブ (Lartruvo) は, PFS で示された治療効果の初期のエビデンスと OS 延長を示唆するデータに基づいて, EMA から条件付き承認を取得しました. 試験の患者数が少ないため, 医薬品の有効性と安全性を検証する実施中の試験で, 追加データを収集することを条件に承認されました[79]. FDA では, 転移性 HER2 陰性乳がん患者の第一選択治療として, ベバシズマブ (Avastin) のパクリタキセル併用治療に迅速承認を与えました. 早期のアウトカムの PFS での治療効果に基づき, PFS の治療効果の検証, 及び OS の治療効果の追加情報の収集が条件づけられ, 承認されました. しかしその後, 臨床試験で十分な有効性が確認されず, 患者に相当な副作用が確認されたため承認は取り消されました[80].

新しい医療技術が代替エンドポイントにより迅速承認されると, その新治療は HTA でも評価されることとなります. NICE やスコットランド医薬品コンソーシアム (Scottish Medicines Consortium, SMC) などの HTA 機関によって行われる償還の決定は, 通常, 新しい医療技術の費用対効果評価に基づいて行われます. その場合, 意思決定モデルに用いる費用や QOL 値を含む多くのパラメータの中で, 特に長期の有効性の推定値が必要になります. 新しい治療法が代替エンドポイントに基づいて承認された場合, 長期の有効性データは, 再評価のためのデータが蓄積するまで利用することができません. 新しい治療法の患者アクセスを促進するための技術評価は, 代替エンドポイントで測定された治療効果の推定値に基づく必要があります. 例えば, NICE 技術評価委員会 (appraisal committee) は, PFS の治療効果の推定値を利用して, 再発又は難治性の慢性リンパ性白血病に対するベネトクラクスのリツキシマブ併用治療を推奨しました[81]. この推奨は, PFS の治療効果が OS の治療効果を代替すると判断されたことを根拠としています.

しかしながら, HTA 機関は代替アウトカムデータの使用には慎重で, ガイドラインで代替エンドポイントを適切に利用することの重要性を強調しています. 実際, 欧州医療技術評価ネットワーク (the European Network for Health Technology Assessment, EUnetHTA) のガイドライン[82]と, NICE の医薬品企業向けの現行の方法ガイド[83]でも言及されています. 代替エンドポイントを用いて費用対効果を予測することにより生じる, 追加の不確実性についても, 十分に検討する必要があります.

NICE 方法ガイド[83]の 4.4.3 節では、適切なアウトカムと代替のアウトカムを特定することの重要性が強調されており、5.7 節 のモデリング方法には、「評価委員会の意思決定プロセスに必要な形式で、利用可能なエビデンスを統合し、臨床的効果及び費用対効果の推定に用いるひとつの重要な枠組み（ほとんどの技術評価にはモデルが必要になります。）」が示されています。ガイドにはモデリングが必要になる可能性が高い状況の説明があります。

「最終的な臨床エンドポイントが利用できない場合に、他のアウトカムによる代替データから生存(年)や HRQoL (health-related quality of life) に対する治療効果を推定する場合、代替エンドポイントから最終的な臨床エンドポイントへの関連を裏付ける証拠を提示する必要がある。また、モデルにおいてその関係がどのように定量化されたのかを説明する必要がある。代替エンドポイントから HRQoL や生存(年)を推定可能という強力なエビデンスが存在する場合、QALY を推定するための代替エンドポイントは最も有用となる。代替エンドポイントを用いるときには常に、代替エンドポイントと HRQoL や生存(年)との関係の不確実性を検討し、定量化する必要がある。」

代替エンドポイントと最終的な臨床アウトカムの関係を評価するために、2 変量メタアナリシスの手法を用いることができます。2 変量メタアナリシスにより、アウトカム間の相関（又はこれらアウトカムで測定された治療効果間の相関）だけでなく、関連する全ての不確実性を考慮に入れることができます。こうした不確実性の意思決定モデルへの反映は、ガイドでも言及されており、相関を考慮（直接又は相関を記述する何らかの関数を通じて）することなく、代替エンドポイントと最終的な臨床アウトカムの関係を定量化することはできません。線形回帰やメタ回帰などの他の方法は、関連する全ての不確実性を考慮しません。例えば、代替エンドポイントの治療効果における不確実性を無視します（補足：通常の回帰分析の仮定では、説明変数は確率変数ではなく与えられた値として定義される）。一方で、2 変量メタアナリシスは、関連する全ての試験内、及び試験間の不確実性を適切に考慮できます。

このように、代替エンドポイントは、医薬品開発プロセスにおいて、試験デザインと評価の両方の段階で非常に重要になります。真のエンドポイントを正確に測定するまでに長い追跡期間が必要とされる環境において、代替エンドポイントは、治療効果を早期に、またはより正確に測定できる場合に特に有用です[68]。特に、前述した抗がん剤の開発の場合、OS が主な関心事である一方で、PFS などの他のアウトカムが早期に利用可能です。特に、新世代の標的治療薬では、真のエンドポイント（OS の場合は死亡するまでの時間）のイベント数が少なく、真のエンドポイントの治療効果に大きな不確実性を生じさせる可能性があります。一方で、PFS に対する治療効果を合理的な精度で推定するために必要な、「増悪（Progression）」のイベント数が十分に多い場合があります。

また PFS が主要な関心事であり、PFS の代替エンドポイントとして腫瘍縮小効果が評価される場合もあります。例えば、再発寛解型多発性硬化症における疾患進行の代替エンドポイントの再発率です[84]。代替エンドポイントは新しい医療技術の認可段階でますます重要な役割を果たしており、HTA の意思決定においても関心が持たれています。

4.1.2 代替エンドポイントの妥当性

代替エンドポイントの候補は、真のエンドポイントの治療効果を予測できるか評価する必要があります。医薬品規制調和国際会議ガイドライン ICH-E9[85]では、「実際には、代替性の証拠の強さ（代替性の基準）は、(i) 生物学的妥当性、(ii) 真のエンドポイントに対する代替エンドポイントの予後的価値の疫学研究での証明、(iii) 代替エンドポイントにおける治療効果が、真のエンドポイントでの治療効果に対応するという臨床試験からの証拠に依存する」と述べられています。

TSD 20 では、これらの基準に加えて、Bucher らの JAMA Users' Guides to the Medical Literature[86]、Lassere らの Outcomes Measures in Rheumatology Clinical Trials (OMERACT) Working Group[87]で発表した同様の代替エンドポイント検証基準、Taylor と Elston[73]、Ciani ら[74]を参照しています。

Taylor と Elston[73]は、医療政策立案者のために設計された、代替エンドポイント評価のための3段階のフレームワークを提案しました（それぞれ、代替エンドポイントの妥当性を裏付ける証拠の利用可能性に依存します）。

- レベル3：生物学的妥当性のみに基づく証拠
- レベル2：個々の患者のレベルで、代替エンドポイントと最終的な臨床エンドポイントの間に強い関連性を示す証拠
- レベル1：多くの無作為化比較試験において、代替エンドポイントを改善する医療技術が真のエンドポイントも改善することを示す証拠

代替エンドポイントの評価では、単一試験のデータに基づいて、個人レベルでの関連性を検討する場合、つまりエビデンスレベル2のみに頼ることは、かなりの批判がありました。Fleming and DeMets [88]は「相関は代替とは異なる (A correlate does not a surrogate make)」という有名な言葉を残しており、Prentice によって開発された代替性の基準について議論しています（補足：基準の詳細は4節冒頭で参照した総説[67]を参照）。本基準では、代替エンドポイントは、真のエンドポイントと相関していること、更に真のエンドポイントでの治療効果に対する完全なる「媒介」でなければなりません（必ず代

替エンドポイントの変化を通じて、真のエンドポイントに影響を与えることが必要)。 Fleming and DeMets は、以下のような幾つかのシナリオを説明しており、それらは、治療法の作用機序に依存し、後者の基準（完全なる「媒介」）は満たされない可能性を指摘しています。 Prentice の代替性の基準の限界は、数多く議論されており[89][90][91]、著者らはこれらの基準を満たしても、代替エンドポイントの治療効果と真のエンドポイントの治療効果との因果関係を保証するものではないことを強調しています。

この因果関係の問題は、単一の試験に基づいて代替エンドポイントが検討される場合、ほとんどの場合、影響を受けます。 Joffe and Greene[89]は、多数の試験から得られたデータに基づくメタアナリシスアプローチは、因果効果（CE：Causal-Effects）ではなく、因果関係（CA：Causal Association）のパラダイムに従っており、（真のエンドポイントに対する代替エンドポイントの効果をモデル化するのではなく）代替エンドポイント候補の治療効果と真のエンドポイントの治療効果との関連性を確立することに基づいていると、議論しています[89]。著者らは、本アプローチは因果効果（CE）パラダイムの制約に縛られないため、代替エンドポイント評価により有用であると指摘しています。因果効果パラダイムでの検討方法の概要は田中ら（2010）[67]を参照してください。因果関係（CA）の研究に対するメタアナリシス的アプローチでは、無作為化から直接得られるものであり、平均因果効果のような治療効果間の関連性を含みます。このアプローチは、多数の試験又はサブグループのデータに基づいており、異質な対比（異質な治療効果）を含む可能性が高く、単一の試験に基づく評価よりも明らかに有利です。単一試験による検証では、ある治療法で確認された関連性が、他の介入でも成立することを保証できません。

4.1.3 代替エンドポイント評価におけるメタアナリシス

代替エンドポイントの代替性を評価するため、様々なメタアナリシス手法が提案されています。代替エンドポイントの候補は、メタアナリシス手法を用いて、様々な集団、且つ様々な治療法で検討された複数の試験を通じて、代替エンドポイントと最終的な臨床エンドポイントとの治療効果間の関連パターンを推定することで、妥当性が証明されます[52][69][70][71][55][72]。多変量メタアナリシス手法は、複数のエンドポイント間の相関を考慮しながら、複数のエンドポイントの平均治療効果を推定するために用いられます[34][36][38][37]。したがって、代替エンドポイントのモデリングに適した方法です[55][72]。例えば、真のエンドポイントの治療効果の不確実性だけでなく、代替エンドポイントの不確実性も考慮に入れるため、（代替エンドポイントの治療効果を、変量効果（確率変数）ではなく、固定効果として扱う）メタ回帰モデルより優れています。代替エンドポイント及び真のエンドポイントに対する治療効果に関する 2 変量メタアナリシスにより、代替エンドポイントの検証と、代替エンドポイントで観察された治療効果

から真のエンドポイントに対する観察されていない治療効果の予測の両方が可能となります。TSD 20 では、（個人レベルではなく）試験レベルでの代替エンドポイントの評価に対して 2 変量メタアナリシス手法を適用することに焦点を当てています。

個人レベルと試験レベルの両方で代替性を評価する IPD モデル（一部の試験では個別患者データが利用可能な場合のモデリング）については、3.1 節で簡単に説明されています。

4.1.4 代替エンドポイント検証のデータ要件

2 変量メタアナリシス手法を利用し、代替エンドポイントを評価するには、両方のアウトカム（代替エンドポイントと真のエンドポイント）の全ての関連する試験データの治療効果と、対応する標準誤差が分析に必要になります。一度に複数の代替エンドポイントを検討する場合、同じ情報を含む（すなわち、候補となる代替エンドポイントに対する治療効果と、真のエンドポイントに対する治療効果と、それぞれの標準誤差）、データセットが必要となります（候補となる代替エンドポイントの数）。データセットの例が TSD20 の付録 B.1 に示されているため、適宜参照して下さい。1 つのアウトカムの治療効果のみを報告する試験は、通常なら分析には含めないでしょう。しかし、これらの試験にて未報告の治療効果はモデルにより予測でき、予測された治療効果は試験間相関の推定に寄与する可能性があるため、それらの試験を分析に含めることもできます（3.3 節参照）。多変量メタアナリシスを使用して、複数の代替エンドポイントを同時に検討することもできます。複数の治療効果（全ての代替エンドポイントと真のエンドポイント）と対応する標準誤差を含むデータが必要になります。

4.6.1 節で説明する交差検証法（クロスバリデーション）の手順を実行する場合、同じ形式のデータが必要となります。

4.6.2 節のように、新しい試験で代替エンドポイントの治療効果から真のエンドポイントの治療効果を予測する場合、クロスバリデーションの手順とは異なり、真のエンドポイントの治療効果は不明であり、治療効果が欠測値としてモデルに含まれます。治療効果を予測する場合、予測値の不確実性を考慮に入れたいため、予測区間も推定する必要があります。そのため、母集団レベルの治療効果の分散について追加の仮定を行う必要があります。したがって、各試験の患者数とともに、対応する母集団の分散と一緒に治療効果データがあると便利です。4.6 節で詳しく説明されています。

通常、系統的（システマティック）レビューを用いて関連する試験が特定されることから、文献レビューの範囲は、例えば HTA 組織による評価中の技術よりもはるかに広く

なる可能性があります。強力な代替エンドポイント（臨床的利益の優れた予測因子）であれば、代替の関係性は、1つの治療法又は1つのサブグループのデータに基づくものではなく、当該疾患領域の全ての試験での全てのサブグループから得られたデータに基づくでしょう。ただし多くの場合、治療法のサブセット、又は部分集団のみが含まれます。治療の種類又はサブグループ間の作用機序の違いが、代替エンドポイント及び真のエンドポイントの治療効果の推定値に、異なる影響を及ぼすことで、代替的关系性の推定に影響を与えてしまう場合に該当します。

4.2 標準的な代替性のモデル(Daniels and Hughes)

Daniels and Hughes[52] [8]により提案されたモデルでは、代替エンドポイント Y_{1i} と真のエンドポイント Y_{2i} の治療効果は、2変量正規分布から生成されると仮定され、代替エンドポイント及び真のエンドポイントに関する真の効果 (δ_{1i} 及び δ_{2i} ; 試験 i の治療効果) を推定します。対応する試験内の標準偏差及び試験間の標準偏差 (σ_{1i} 及び σ_{2i}) , 試験内の2変量の相関 ρ_{wi}^{12} を用いて、以下のとおりに表すことができます。

$$\begin{pmatrix} Y_{1i} \\ Y_{2i} \end{pmatrix} \sim N \left(\begin{pmatrix} \delta_{1i} \\ \delta_{2i} \end{pmatrix}, \begin{pmatrix} \sigma_{1i}^2 & \sigma_{1i}\sigma_{2i}\rho_{wi}^{12} \\ \sigma_{1i}\sigma_{2i}\rho_{wi}^{12} & \sigma_{2i}^2 \end{pmatrix} \right)$$

$$\delta_{2i} | \delta_{1i} \sim N(\lambda_0 + \lambda_1 \delta_{1i}, \psi_2^2) \quad (7)$$

ここで、代替エンドポイントの真の効果 δ_{1i} は、試験に固有の固定効果であり、真のエンドポイントの真の効果 δ_{2i} と線形関係を有すると仮定します。全てのパラメータに事前分布が与えられます。例えば、固定効果には $\delta_{1i} \sim N(0, 1000)$, 及び回帰パラメータには、 $\lambda_0 \sim N(0, 1000)$, $\lambda_1 \sim N(0, 1000)$, $\psi_2 \sim Unif(0, 2)$ 等です。

Daniels and Hughes は強力な代替性として、 $\lambda_0 = 0$ (代替エンドポイントの治療効果がない場合には、真のエンドポイントの治療効果がないとする条件) , 及び $\lambda_1 \neq 0$ (代替エンドポイント及び真のエンドポイントとの治療効果間関係性を確立させる条件) を挙げました。また、条件付分散 ψ_2^2 は関連の強さを測定しており、完全な代替性が成立する場合には、0と等しくなります。代替性の関係の強さを測るこれら3基準を「代替性の基準」として参照します(4.1.2節参照)。

4.3 積・正規分布配合における BRMA

Daniels and Hughes[52]のモデルと同様に、5章で説明する積・正規分布の定式化の BRMA (bivariate random-effects meta-analysis) モデルは、代替エンドポイント評価にも利用できます。2つの方法の主な違いは、BRMA モデルでは、代替エンドポイントの治療効果は変量効果（試験間の交換可能性）が仮定され $\delta_{ii} \sim N(\eta_1, \psi_1^2)$ と表される一方、Daniels and Hughes では固定効果と仮定されていることです。交換可能性の仮定が妥当である場合、BRMA モデルは、（両方のアウトカムについて）全試験を通じて借用できる情報が増えるため、より正確な予測につながる可能性があります。ただし、Daniels and Hughes によって説明されているように、変量効果の分布の選択は困難であり、実薬対照治療と（単なる）対照治療の順序に依存することがあります。例えば、同じ対照群に対して、2つのクラスの治療法が大きく異なる治療効果を有する場合、それらの対照群に対する治療効果の分布が二峰性になってしまう等、複雑になる可能性があります。容易にモデルを誤特定してしまう可能性があります。Bujkiewicz らによるシミュレーション研究[55]によると、このように交換可能性の仮定が成り立たない場合、5章で紹介する BRMA PNF モデルから得られた予測が偏る可能性があります。

本モデルの代替性の基準は、前節の Daniels and Hughes によるモデルと同じです。

4.4 標準形式の BRMA

5章で紹介する標準形式の BRMA は、代替エンドポイントの評価にも利用できます。この場合、代替パターンの強度の測定尺度は、試験間の相関 ρ です。代替性が完全である場合、相関は $\rho = \pm 1$ です。5章で紹介する BRMA PNF と同様に、本モデルも代替エンドポイントの変量効果を仮定していることに注意して下さい。

4.5 文献の他のモデルとの関係

代替エンドポイントの評価では、各方法の利点と潜在的な限界のバランスを考慮する必要があります。モデルの適合性を検討するには、感度分析を実施することを推奨します。感度分析は、DIC (Deviance Information Criteria) [92]を使用してモデルを比較すること、及び4.6.1節でより詳細に説明するクロスバリデーションによって実施できます。

4.2節から4.4節で説明したモデルパラメータ、及びそれらの関係を表1に要約します。BRMA モデルのパラメータ、平均 d_j 及び異質性 τ_i 、及び2つのアウトカムの真の効果の

試験間の相関 ρ は、BRMA PNF モデルのパラメータ、及び対応する代替性基準に直接関連します。傾きと条件付き分散は以下の式 (8) と式 (9) となります。

$$\lambda_1 = \rho \frac{\tau_2}{\tau_1} \quad (8)$$

$$\psi_2^2 = \tau_2^2 - \lambda_1^2 \tau_1^2 = \tau_2^2(1 - \rho^2) \quad (9)$$

傾きは $\lambda_1 = (d_2 - \lambda_0)/d_1$ でもあることから（回帰直線の図を描くことにより簡単に確認できます。補足：切片 λ_0 及び傾き λ_1 の直線が点 (d_1, d_2) を通る回帰直線より）、BRMA のパラメータに関して以下の切片を持ちます（補足：(8)の右辺を代入し、 λ_0 に関して解く）。

$$\lambda_0 = d_2 - d_1 \rho \tau_2 / \tau_1 \quad (10)$$

BRMA PNF モデル（Daniels and Hughes[52]によるモデルと同様に）では、条件付き分散が 0 ($\psi_2^2 = 0$) である場合、代替性の関係性は完全です（Daniels and Hughes, Bujkiewiczら[52][55]）。式 (9) から、 $\rho^2 = 1 - \frac{\psi_2^2}{\tau_2^2}$ であり、もし $\psi_2^2 = 0$ であれば、相関は $\rho = \pm 1$ となります。したがって 2 つの BRMA モデルのための完全な代替性基準は同等であることがわかります。本基準は、Daniels and Hughes モデルの代替性基準と同じ意味を持っていますが、対応するパラメータは BRMA モデルのパラメータに直接関係していません。これは、Daniels and Hughes モデルの固定効果の仮定によるものであり、つまり異質性のパラメータや平均のパラメータが存在しません。

表 1：代替エンドポイントに対するモデルパラメータとそれらの関係性

Daniels and Hughes	BRMA PNF	BRMA
λ_0	λ_0	ρ
λ_1	λ_1	τ_1
ψ_2^2	ψ_1^2	τ_2
	ψ_2^2	d_1
	η_1	d_2
	$\eta_1 = d_1$	$d_1 = \eta_1$
	$\lambda_1 = \rho \tau_2 / \tau_1$	* $d_2 = \lambda_0 + \lambda_1 \eta_1$
	$\lambda_0 = d_2 - d_1 \rho \tau_2 / \tau_1$	$\rho = \lambda_1 \tau_1 / \tau_2$
	$\psi_1 = \tau_1$	$\tau_1 = \psi_1$
	$\psi_2^2 = \tau_2^2(1 - \rho^2)$	$\tau_2^2 = \psi_2^2 + \lambda_1^2 \psi_1^2$

$$\rho^2 = \lambda_1^2 \psi_1^2 / (\psi_2^2 + \lambda_1^2 \psi_1^2)$$

λ_0 - 切片, λ_1 - 傾き, ψ_2^2 - 条件付き分散, ρ - 試験間の相関, η_1, d_1, d_2 - 平均値, τ_1, τ_2 - 試験間の異質性パラメータ

* 効果を中心 ($d_2 = \lambda_0$) で推定した場合 (5章参照)

代替性基準は ρ^2 として表現でき ($\rho^2=1$, 完全な代替性), 試験レベルを調整した R-squared と呼ぶ研究者もいます (Burzykowski ら[70]及び Renfro ら[71]) .

[出典:NICE DSU TSD 20(<https://nicedsu.sites.sheffield.ac.uk/tsds/completed-or-in-progress>)]

4.6 検証と予測

4.6.1 クロスバリデーションの手順

代替エンドポイントを評価するには, まずは全てのデータのメタアナリシスから代替性基準を得ます. つまり, Daniels and Hughes[52]によるモデル, 及び BRMA PNF モデルの回帰係数, 又は標準形式 BRMA を用いる場合は, 試験間相関(又は調整された R-squared)です. 次に, 真のエンドポイントに対する治療効果の予測値(及び対応する予測区間)を, 1 試験ずつ除くクロスバリデーション法で推定した, (複数の) 予測値と比較します. 一度に1つの試験を除くため, 当該試験 i の真のエンドポイントに対する治療効果の推定値 Y_{2i} が除外され(そして, MAR として扱われ), メタアナリシスの残り全ての試験での両アウトカムデータの条件付きのもと, 代替エンドポイントの治療効果から真のエンドポイントの治療効果を予測します. 平均予測効果は, 真のエンドポイントの治療効果の平均予測 $\hat{\delta}_{2i}$ に等しく, MCMC シミュレーションにより予測され, 分散が $\sigma_{2i}^2 + \text{var}(\hat{\delta}_{2i} | Y_{1i}, \sigma_{1i}, Y_{1(-i)}, Y_{2(-i)})$ となります. ここで, $Y_{1(-i)}$ 及び $Y_{2(-i)}$ は, バリデーション試験 i を除いた残り試験のデータに基づいたデータであることを示す. 妥当な代替エンドポイントの場合, この分散に基づいて構築された予測区間は, 95% 予測区間を算出する場合, 少なくとも 95% の試験で観測された推定値が含まれている必要があります.

欠測データが2変量メタアナリシスに存在する場合, ベイズ流の枠組みでは, 治療効果の欠測は, MCMC シミュレーションを利用してメタアナリシスモデルから直接予測されます. これを実装するには, 欠測データに事前分布を与える必要があります. 治療効果の統合モデルでは, 試験 i の欠測アウトカム j の治療効果 Y_{ji} の欠測値は既に考慮されています. ただし, 対応するアウトカムの標準誤差 σ_{ji} に対しても事前分布を与える必要があります. クロスバリデーションでは, 代替エンドポイントか真のエンドポイントの

うち、1つの治療効果のみが欠測しています。標準誤差の事前分布として、例えば、 $\sigma_{ji} \sim Unif(0.0001, 15)$ を与えます。

2.1.3 節では、併合した治療効果 (pooled effect) を推定するための2変量メタアナリシスの利用について説明し、欠測した標準誤差の推定方法として、母集団の分散の交換可能性を許容するモデルが紹介されました。ただし、クロスバリデーションの手順の中では、予測区間を過大評価又は過小評価する可能性があるため、そのモデルは適切なアプローチではありません。しかしながら、クロスバリデーションでは欠測の標準誤差を推定する必要がないため、欠測の標準誤差に独立した事前分布を設定することで十分です。(本節の最初の段落の式のように) 予測区間を構成するとき、実際の標準誤差を使用します(クロスバリデーションの設定では未報告として取り扱いますが、標準誤差は実際には既知です)。しかしながら、代替エンドポイントの治療効果に関するデータのみが利用可能で、真のエンドポイントデータがない新しい試験の真のエンドポイントの治療効果を予測したい状況では、予測効果の不確実性を適切に説明するため、標準誤差を推定する必要があります。

4.6.2 新しい試験における代替エンドポイントの治療効果による、真のエンドポイントに対する治療効果の予測

クロスバリデーションが成功し、真のエンドポイントの代替エンドポイントとして利用できると結論付けたら、代替エンドポイントの治療効果のみが報告されている新しい試験で、真のエンドポイントの治療効果を予測ができます。予測には2変量メタアナリシスモデル、すなわち、4.2 節で説明した Daniels and Hughes モデル[52]、又は 4.3 節と 4.4 節で説明した BRMA モデルを利用します。

繰り返しになりますが、Daniels and Hughes モデルと BRMA モデルの主な違いは、前者は代替エンドポイントの治療効果に対して固定効果(独立した事前分布)を、後者では変量効果を仮定していることです。代替エンドポイントの変量効果(すなわち、真の効果 δ_{ij})の正規性の仮定が満たされていない場合(例えば、代替エンドポイントの推定効果ヒストグラムが、二峰性になる等、明らかな正規性からの逸脱が認められる場合)、Daniels and Hughes モデルがより適切になります。ただし、正規性の仮定が妥当である場合は、変量効果の仮定は、情報を追加的に借用できるため、BRMA モデルの方が、より正確な予測が得られる可能性があります。

データが欠測した場合の BRMA モデルと同様に、新しい試験の未報告の真のエンドポイントの治療効果と対応する標準誤差を、欠測としてコード化します。ただし、データが欠測した場合の BRMA モデルと同様に、真のエンドポイントに対する予測効果の信

用区間を推定できるように、母分散の交換可能性を仮定します。その場合、両アウトカムの治療効果と症例数に対応する(標準誤差ではなく)母分散データが必要となります。詳細は3節を参照して下さい。代替エンドポイントの治療効果、未報告の真のエンドポイントの治療効果、及び対応する母集団レベルの分散は、両アウトカムを報告している他の試験データと同じモデルに従うと仮定することで、新しい試験の未報告の真のエンドポイントの治療効果と対応する分散は、MCMCシミュレーションにより(全てのデータを条件とする)モデルから予測されます。新しい試験の真のエンドポイントの予測効果は、新しい試験 k の Y_{2k} の平均値と信用区間として得られます。

次に、真のエンドポイントは報告されていないが、代替エンドポイントの治療効果が報告されている新しい試験が複数ある場合を考えます。標準的な(単変量)メタアナリシス等を利用し、平均予測効果を個別又は同時に取得し、意思決定の枠組みで利用できる個別の予測推定値として扱うことができ、複数の新しい試験の真のエンドポイントの治療効果を予測できます。特に、新しい試験が様々な治療法を検討している場合、各試験の個別の予測効果が注目される可能性が高いです。

4.7 例:再発寛解型多発性硬化症における代替性の妥当性評価

4.2節から4.6節で説明したモデリング手法を説明するため、TSD 20では、再発寛解型多発性硬化症の例に適用していますが、本書では割愛します。実際に適用する場合は、TSD 20の3.7節を参照して下さい。

4.8 代替性の基準、他の代替エンドポイントモデル及びさらなる研究の議論

上述したように、試験間相関が ± 1 の場合は完全な代替性を示します。現実的には、予測を行うのに適した代替エンドポイントを検討するために、相関の大きさがどの程度であるべきかを定量化することは困難です。一部の著者は、代替性を実証するには高度な相関関連が必要であると主張しています。例えば、Lassereらは、「Biomarker-Surrogacy Evaluation Schema」にて、試験間相関係数の二乗(いわゆる調整済みR-Squared)が0.6(注:相関係数0.77)を超える高い関連性を定義しました[93]。またドイツのIQWiG(Institute for Quality and Efficiency in Health Care:医療の質と効率性評価機関)では、相関係数の95%信頼区間の下限が0.85を超える高い相関が必要としています[94]。他の著者は、代替エンドポイントが臨床的利益を予測し得るかの決定は、関係性の強さと新規治療法の有効性に対して行われる決定(例えば規制の目的)の必要性とのバランスに基づくべきと強調しています[95]。さらに、代替関係の強さ(又は弱さ)は、真のエン

ドポイントの治療効果の予測区間の幅の広さを見れば明白です。相関が小さいと区間幅が広くなり、そのような予測に基づいて行われた規制や臨床の決定では不確実性が高まります。すなわち、相関係数に関する基準は必要なく、代わりに予測結果を検討するだけで良いと、筆者達は述べています[43]。治療効果の（不確実性ととも）予測された推定値は、医療経済性の意思決定に利用できます。予測の質の評価は、4.6.1 節で説明されたクロスバリデーションにより行います。

本書では、代替エンドポイントの評価では、4.1.2 節（Taylor and Elston[73]）で説明したエビデンスレベル 1 である要約データに対するメタアナリシスに焦点を当てました。これは規制上の意思決定で最も適切となります（補足：英国 NICE の場合です）。IPD が使用可能な場合、代替エンドポイントは、個人レベルと試験レベル（エビデンスレベル 2 及びレベル 1）の両方で評価できます。

Buyse らの方法[69]は、Prentice[96]によって開発された方法をメタアナリシスの枠組みに拡張することにより、群レベルで代替エンドポイントをモデル化するために開発されました。候補となる代替エンドポイントの妥当性を、両方のエビデンスレベルで同時に検討するため、個別レベルの R-Squared と、試験レベルの R-Squared の 2 つの尺度を用い、混合効果モデルの枠組みを開発しました。本方法は頻度論的アプローチです。本方法の様々な拡張が提案されており[68]、例えば、Burzykowki ら[70]は事象までの時間（time-to-event）、Renfro ら[71]はベイズ流の枠組みで拡張しています。

4.9 その他の発展的なモデル

代替関係の程度は、治療法又は治療クラスの作用機序にも関係している場合もあります。そのような場合、従来の二変量メタアナリシスによる方法では、同質とみなせるグループを対象に検討することがあります。しかしながら、データソースが少なくなるといった問題があり、これに対応するために、Bujkiewicz ら[43]及び Papanikos ら[97]によって、新しい方法が開発され、適用例やシミュレーション研究による評価も報告されています（それぞれの筆頭著者は TSD 20 の執筆メンバーでもあります）。これらの方法について、TSD 20 でも少しだけ触れられていますので、本節では、もう少しだけ詳しくまとめました。なお、Bujkiewicz らの提案モデルの一部と Papanikos らの提案モデルについては、論文の補足資料に BUGS コードが記載されています。

Bujkiewicz ら[43]は代替エンドポイント評価のための二変量メタアナリシス法を拡張し、二変量ネットワークメタアナリシス法を提案しています。提案法では、治療比較ごとに別々の多変量分布に従うと仮定します。著者らの提案する方法のうち基本とな

る考えのモデルを以下に示します。この試験単位の代替性の関係をモデル化したものに、制約や仮定をおいた複数のモデルが提案されていますので、詳細は論文をご参照ください。

$$\begin{pmatrix} Y_{1kli} \\ Y_{2kli} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \mu_{1kli} \\ \mu_{2kli} \end{pmatrix}, \Sigma_i \right\}, \Sigma_i = \begin{pmatrix} \sigma_{1kli}^2 & \sigma_{1kli}\sigma_{2kli}\rho_{wkli} \\ \sigma_{1kli}\sigma_{2kli}\rho_{wkli} & \sigma_{2kli}^2 \end{pmatrix}$$

$$\begin{pmatrix} \mu_{1kli} \\ \mu_{2kli} \end{pmatrix} \sim N \left\{ \begin{pmatrix} d_{1kl} \\ d_{2kl} \end{pmatrix}, \begin{pmatrix} \tau_{1kl}^2 & \tau_{1kl}\tau_{2kl}\rho_{1kl,2kl} \\ \tau_{1kl}\tau_{2kl}\rho_{1kl,2kl} & \tau_{2kl}^2 \end{pmatrix} \right\},$$

Y_{jki} : 試験 i におけるアウトカム $j=1, 2$ に対する治療効果の差 (治療 k 対 l)

μ_{jki} : 試験 i におけるアウトカム j への真の治療効果 (治療 k 対 l)

d_{jki} : アウトカム j への平均治療効果 (治療 k 対 l)

完全に代替関係となる場合, $\rho_{1kl,2kl} = \pm 1$ and $\mu_{1kli} = 0 \Leftrightarrow \mu_{2kli} = 0$.

このモデルにより、既存の治療法の新しい環境での有効性や新しい治療法を検討する新規試験において、代替エンドポイントで観察された効果から、真のエンドポイントに対する治療効果を予測することができます。

ここで、新しい治療法を評価する新規の試験によって、代替エンドポイントの効果から真のエンドポイントに対する治療効果を予測したい場合を考えてみます。新しい治療法の試験は1つだけであり、代替エンドポイントのみのデータであることも予想されます。この場合は、事前分布のみに基づいて、新しい治療についての真のエンドポイントに対する平均効果を推定することになってしまいます。そこで著者らは、前述した試験単位での代替関係をモデル化することに加えて、治療間の交換可能性 (類似性) を追加で仮定することによって、治療単位の代替性もモデル化するという手法の拡張も提案しています。この拡張された手法により、新しい試験、新しい治療に対する真のエンドポイントに対する治療効果を予測することができます。しかしながら、例えば、異なる患者集団で異なる作用機序の治療法について評価する場合には、この仮定が満たされない可能性もあるため、慎重に検討する必要があります。

Papanikos ら[97]による方法では、治療クラス間の代替関係の情報を借用することを可能とした方法を提案しています。著者らは2つの方法を説明しており、1つ目の方法は、代替エンドポイントを評価するための標準的な Daniels and Hughes モデル (4.2 節) を、すべての治療クラスにおける代替関係に完全な交換可能性を仮定した階層的なメタアナリシスモデルに拡張することで、クラス間の情報の借用を容易にしています。2つ目の方法では、この仮定を緩和し、治療クラス間の代替関係に部分的な交換可能性を仮定し、完全な交換可能性を仮定するには合理的でないときにより柔軟とみ

なせます。なお、提案法では治療クラス内の薬剤を区別していませんが、治療クラス内の異なる治療法を考慮して、さらなる階層を追加することもモデルの拡張方法として挙げられてはいます。ただし、多くの試験が必要になるだろうとも述べられています。以下に二つのモデル式の概要を示します。

1つ目のモデル：Hierarchical model with full exchangeability (F-EX) model

治療クラス o の試験 i の代替エンドポイントおよび真のエンドポイントの治療効果 Y_{1io} , Y_{2io} に対するモデル

$$\begin{aligned} \begin{pmatrix} Y_{1io} \\ Y_{2io} \end{pmatrix} &\sim N\left(\begin{pmatrix} \mu_{1io} \\ \mu_{2io} \end{pmatrix}, \begin{pmatrix} \sigma_{1io}^2 & \sigma_{1io}\sigma_{2io}\rho_{wio} \\ \sigma_{1io}\sigma_{2io}\rho_{wio} & \sigma_{2io}^2 \end{pmatrix}\right) \\ \mu_{2io}|\mu_{1io} &\sim N(\lambda_{0o} + \lambda_{1o}\mu_{1io}, \psi_o^2) \\ \lambda_{0j} &\sim N(\beta_0, \xi_0^2), \lambda_{1o} \sim N(\beta_1, \xi_1^2). \end{aligned}$$

σ_{1io} , σ_{2io} , and ρ_{wio} : 治療クラス o の試験 i の試験内 SD および試験内相関

λ_{0o} は治療クラス o の切片, λ_{1o} は代替エンドポイントへの治療効果と治療クラス o 内の真のエンドポイントとの関係 (傾き) .

2つ目モデル：Hierarchical model with partial exchangeability (P-EX) model

F-EX model の λ_{1o} について拡張し, 治療クラス間の代替関係に部分的な交換可能性を仮定したモデルは以下になります。

$$\lambda_{1o} = \begin{cases} \lambda_{1o} \sim N(\beta_1, \xi_1^2) & \text{if } p_j = 1 \\ \lambda_{1o} \sim N(0, b) & \text{if } p_j = 0 \end{cases}$$

ここで, 治療クラス o について, ベルヌーイ分布に従う p_o を用い, $p_o=1$ の場合は他の治療クラスの傾きと共通, つまり, すべての治療クラスに対して $p_o=1$ のときに F-EX model と同等になるようにしています. $p_o=0$ の場合は他の治療クラスの傾きとは交換可能でないことを表しています. ξ_1^2 と b は, それぞれのシナリオにおける分散に対応します. 著者らの実装例では $\lambda_{1o} \sim N(0, 1000)$ としていました. p_o の事後平均を算出することで, 交換可能性の程度をデータから推測することができます. なお, p_o のパラメータ π_o の事前分布には, $\pi_o \sim \text{Beta}(1, 1)$ などとすることも可能ですが, MCMC が収束するのに長い連鎖(chains)が必要です.

従来の二変量メタアナリシスによる方法を用いて, 同質とみなせるグループ, 例えば治療クラスごとに解析する方法 (サブグループ解析) はシンプルな手法であり, 各治療クラスに十分なデータがある場合には良好な結果が得られますが, 治療クラス内の

データが少ししかない場合も想定されます。著者らによる提案法（F-EX model, P-EX model）と前述のサブグループ解析を比較するシミュレーション研究の結果によると、提案手法は代替関係を記述するパラメータの事後平均の平均絶対誤差や二乗平均平方根誤差を（平均的に）減少させるとともに、真のエンドポイントの効果の予測を改善することができました。特に P-EX モデルは、他の治療クラスの傾きと明らかに異なる治療クラスが存在する状況において、他の方法より好ましい結果が得られました。しかしながら、実際の場面では、十分な数の治療クラスを持つデータセットを見つけることが困難な場合があり、治療クラスの数が少ないと、提案法のパフォーマンスに影響があります。そのため潜在的な限界を常に念頭に置いておく必要があるとも述べられています。

論文中では、提案手法をさらに発展させることについても挙げられており、その中には、前述の Bujkiewicz らによるネットワークメタアナリシスの手法と組み合わせることについても記載されています。階層モデルを拡張し、複数の代替エンドポイント（または、複数の時点で報告された同じ種類の代替エンドポイント）を真のエンドポイントに対する治療効果の予測因子としてモデル化する構想が考えられているようです。

本節にて紹介した手法は、本報告書の発出時点では新しい方法であり、実際に使用された例はないようです。一方で、これらの手法の筆頭著者の方々は TSD 20 の作成に関与していますので、今後はこのような方法を用いた評価が求められるかもしれません。しかしながら、これらの手法は複雑な仮定を前提とした複雑なモデルを用います。そのため、それぞれの仮定に合理的な根拠はあるのか慎重に検討するとともに、分析前協議などで合意することも必要でしょう。望ましくは、シナリオ分析、感度分析などにより、結論に与える影響もみておきたいところです。実際に適用した際の課題や手法に関するさらなる研究の報告を今後期待したいと思います。

4.10 推奨事項

TSD 20 で述べたエビデンス統合手法は、代替エンドポイントの妥当性検証や（意思決定モデリングフレームワークで使用することが可能な）真のエンドポイントに対する相対的な治療効果の予測という目的には適していますが、意思決定モデリングにおいて代替エンドポイントが果たすべき役割のすべての側面を捉えてはいません。最終アウトカムに関するデータが十分でない場合、生存時間曲線の外挿を行う必要があるでしょう。このような外挿をどのように行うか、またどのようなデータ（例えば、代替エンドポイントに基づく RCT から、あるいは観察研究データのような他の情報源から）を用いる

かは、データの入手可能性と十分に依存します。代替エンドポイントの使用との関連性において、自然史モデルでベースラインの生存時間パターンを予測する最適なアプローチを調査する必要があります。

また、2020年にNICEから発行された、CHTE2020 sources and synthesis of evidence DSU 報告書[98]にて、代替エンドポイントに関する推奨事項がまとめられていますので、以下にお示しします。

- 代替エンドポイントを評価に用いる場合には、（利用可能なデータ及びモデルの仮定を適切に考慮した上で）メタアナリシスのモデリング手法を用いることを推奨します。その詳細はCHTE2020 sources and synthesis of evidence[98]及び本TSD 20を参照してください。
- 過去に公表されたモデルがあれば、代替エンドポイントの妥当性を評価するため、あるいは予測するために使用できます。ただし、そのモデルが同一又は関連する設定（患者集団、すべての治療法又は評価中の新技術に関連する治療クラス内で代替の関係性が成立することを示唆する幅広い治療選択肢）で開発され、予測を行う際にすべての関連する不確実性を考慮できるように、十分詳細に報告されている場合にのみ使用できます。新しい代替エンドポイントを調査する場合や、新しいタイプの治療法/治療ライン（あるいは新しい集団）を検討する場合、新しい状況での代替性が成立しない可能性がある場合には、上記が困難となることがあります。過去のモデルが異なる環境で収集されたデータに基づいている場合、適切なメタアナリシスを用いて新しいモデルを開発することが推奨されます。これには、治療クラス間の作用機序の違い、又はファーストインクラスの薬剤の場合は、それらを反映したネットワークメタアナリシスや階層的手法が推奨されます。
- 代替エンドポイントの妥当性に関するエビデンスが限られているとき、より十分なデータが入手可能になり、調査中の医療技術が再評価され、費用対効果の結果が確認されることを条件に、（医療技術の使用に関して）勧告する可能性があります。これは、特にがん治療薬に推奨されることです。このアプローチは、EMAが代替エンドポイントに基づいて行う販売許可決定にも推奨され、がん治療薬のHTAの意思決定にも適用可能です。NICEとEMAとで協調することで、プロセスが改善される可能性があります（補足：本邦でも、新薬審査と費用対効果評価が協調することで、重複や齟齬が解消され、効率化が期待できると考えます）。

5. 多変量変量効果メタアナリシスのモデル

TSD 20 の 2 章, 4 章, 5 章で紹介されているベイズ流の多変量変量効果メタアナリシスに用いる種々の統計モデルを表 1~5 に示しました. 全てのモデルは, 3 章の式 (1), (2) の拡張となっています.

表 2 二変量変量効果メタアナリシスモデル (BRMA)

モデルの名称	二変量変量効果メタアナリシスモデル (BRMA)
TSD のセクション	2.1 BRMA in a Bayesian framework with considerations of appropriate prior distributions 2.2 BRMA in the product normal formulation (PNF)
試験レベルのモデル	<p>治療効果を表す 2 つの変数 Y_{1i} と Y_{2i} が相関のある 2 変量正規分布に従うと仮定した BRMA モデルは, van Houwelingen ら[34]と Riley ら[36]の記法を用いて次のように記述できます.</p> $\begin{pmatrix} Y_{1i} \\ Y_{2i} \end{pmatrix} = N \left(\begin{bmatrix} \delta_{1i} \\ \delta_{2i} \end{bmatrix}, \begin{bmatrix} \sigma_{1i}^2 & \sigma_{1i}\sigma_{2i}\rho_{wi} \\ \sigma_{1i}\sigma_{2i}\rho_{wi} & \sigma_{2i}^2 \end{bmatrix} \right)$ <p>添字 i が試験を表すとして, 試験レベルの治療効果 (注, 特に断らない限り治療効果は治療間の差/比など) を表す 2 つの変数 (Y_{1i}, Y_{2i}) は, 平均 (δ_{1i}, δ_{2i}), 試験内分散 ($\sigma_{1i}^2, \sigma_{2i}^2$), 試験内相関係数 ρ_{wi} の 2 変量正規分布に従います. この階層的なモデルでは, 試験レベルの効果は交換可能 (exchangeable) と仮定します. つまり, δ_{1i} と δ_{2i} は, 共通の 2 変量正規分布に従います.</p>
母数レベルのモデル① 標準形	$\begin{pmatrix} \delta_{1i} \\ \delta_{2i} \end{pmatrix} = N \left(\begin{bmatrix} d_1 \\ d_2 \end{bmatrix}, \begin{bmatrix} \tau_1^2 & \tau_1\tau_2\rho \\ \tau_1\tau_2\rho & \tau_2^2 \end{bmatrix} \right)$ <p>δ_{1i} と δ_{2i} が従う 2 変量正規分布の平均を (d_1, d_2), 試験間分散を (τ_1, τ_2), 試験間相関係数を ρ とします. 各試験の治療効果 (δ_{1i}, δ_{2i}) の推定値は, 他の試験の情報を借用したベイズ流の縮小推定 ("shrunken" estimates) になります[44].</p>
母数レベルのモデル② PNF による定式化	<p>BRMA モデルは, Bujkiewicz らによって提案され, 代替エンドポイントのモデルに拡張されたように[37], 別の形でも定式化できます. 具体的には, 試験レベルのモデルは同じで, 母数レベルのモデルを PNF, すなわち, 単変量の条件付き正規分布の積で表します[99][37].</p> $\begin{cases} \delta_{1i} \sim N(\eta_{1i}, \psi_1^2) \\ \delta_{2i} \delta_{1i} \sim N(\eta_{2i}, \psi_2^2) \\ \eta_{2i} = \lambda_0 + \lambda_1 \delta_{1i} \end{cases}$

	<p>モデル式の全てのパラメータに独立な無情報事前分布を与えるのではなく、試験レベルのモデルのパラメータと共分散行列の要素の関係を、パラメータ相互の関係として保持するように導出します</p> $\begin{cases} \psi_1^2 = \tau_1^2 \\ \psi_2^2 = \tau_2^2 \\ \lambda_1 = \frac{\tau_2}{\tau_1} \rho \end{cases}$ <p>これらの関係のもとで、試験間の標準偏差と相関に事前分布を与えることができます。パラメータ値の妥当な範囲が既知であったり、外部情報から得ることができる場合には、全てのパラメータに独立な無情報事前分布を与えることと比較すると、推定は容易になります。例えば、試験間の相関に $\rho \sim \text{Unif}(-1, 1)$、アウトカム $j=1, 2$ の試験間の標準偏差に $\tau_j \sim \text{Unif}(0, 2)$ と事前分布を与えると、関係式から母数レベルのパラメータ $\lambda_1, \psi_1, \psi_2$ に、暗黙に事前分布を与えることになります。残りのパラメータには、あいまいな事前分布、$\eta_1 \sim N(0, 1000)$、$\lambda_0 \sim N(0, 1000)$ を与えます。</p> <p>この BRMA PNF モデルは、全ての事前分布が同じで、ふたつのアウトカムに全てに欠測がなくデータが利用可能な場合、標準形の BRMA モデルと同等です。しかし、データに欠測値がある場合、2つのモデルの結果は異なり、BRMA PNF モデルの結果はアウトカムの順序にも依存します。</p> <p>BRMA PNF モデルは、代替エンドポイントを評価する際に特に役立ちます。PNF によるモデル化は、アウトカムのペアに対して治療効果間の条件付き独立性を仮定して、パラメータ数を減らし、モデルを簡略化するため、パラメータ数が多く潜在的に多数の試験のデータが必要となるような高次元のモデルにも有用です。</p>
TSD の数値例	<p>2.4 Example: rheumatoid arthritis</p> <p>関節リウマチ患者の二次治療に使用される抗 TNF-α 阻害剤の系統的レビューとメタアナリシス[100]のデータを用います。アウトカムは、Health Assessment Questionnaire (HAQ) と疾患活動性スコア (DAS-28)、米国リウマチ学会 (ACR) の有効性基準、BRMA では、HAQ と DAS-28 を用います。</p> <p>注意すべき点は、この数値例の全ての試験は単群デザインであることです (治療効果はベースラインからの変化)。リウマチの事例を採用した理由は、扱いやすいことと、同じ手法を治療効果のデータ (例えば、対数オッズ比で表された治療効果及びその標準誤差のデータ) に</p>

	も適用可能だからです。しかしながら、単群試験のベースラインからの変化を用いるアプローチは、治療効果の推定値を得るための理想的な方法として推奨されていません。
WinBUGS code	A.1 WinBUGS code for BRMA in the standard form A.2 WinBUGS code for BRMA in product normal formulation

表 3 三変量変量効果メタアナリシスモデル (TRMA)

モデルの名称	三変量変量効果メタアナリシスモデル (TRMA)
TSD のセクション	4.1 MRMA in the standard form 4.2.1 TRMA PNF unstructured model 4.2.2 TRMA PNF structured model
試験レベルのモデル	<p>多変量メタアナリシスは、BRMA の PNF による定式化を多変量に拡張することでモデル化できます。このアプローチには多くの利点があります。共分散行列の全体に事前分布を与えるのとは対照的に、PNF では試験間共分散行列の全ての要素の事前分布を直接制御することができます。また、PNF はアウトカム間の関連パターンを記述するので、複数の代替エンドポイントをモデル化する際にも有用です。</p> <p>ここでは、簡単のため、三変量変量効果メタアナリシス (trivariate random-effect meta-analysis, TRMA)、つまり 3 つの相関するアウトカムに対する治療効果のメタアナリシスに焦点を当て、任意の数のアウトカムに対する相関のある治療効果の MRMA の PNF によるモデル (MRMA PNF) は TSD 20 付録 C.7 に説明があります。</p> $\begin{pmatrix} Y_{1i} \\ Y_{2i} \\ Y_{3i} \end{pmatrix} = N \left(\begin{bmatrix} \delta_{1i} \\ \delta_{2i} \\ \delta_{3i} \end{bmatrix}, \begin{bmatrix} \sigma_{1i}^2 & \sigma_{1i}\sigma_{2i}\rho_{wi}^{12} & \sigma_{1i}\sigma_{3i}\rho_{wi}^{13} \\ \sigma_{2i}\sigma_{1i}\rho_{wi}^{12} & \sigma_{2i}^2 & \sigma_{2i}\sigma_{3i}\rho_{wi}^{23} \\ \sigma_{3i}\sigma_{1i}\rho_{wi}^{13} & \sigma_{3i}\sigma_{2i}\rho_{wi}^{23} & \sigma_{3i}^2 \end{bmatrix} \right)$
母数レベルのモデル① 標準形	$\begin{pmatrix} \delta_{1i} \\ \delta_{2i} \\ \delta_{3i} \end{pmatrix} = N \left(\begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}, \begin{bmatrix} \tau_1^2 & \tau_1\tau_2\rho^{12} & \tau_1\tau_3\rho^{13} \\ \tau_2\tau_1\rho^{12} & \tau_2^2 & \tau_2\tau_3\rho^{23} \\ \tau_3\tau_1\rho^{13} & \tau_3\tau_2\rho^{23} & \tau_3^2 \end{bmatrix} \right)$ <p>分散共分散行列には、分散>0, -1<相関係数<1 という制約があるため、アウトカムが 2 よりも大きい場合には、この条件を満たすように、すなわち分散共分散行列が半正定値行列となるように事前分布を与える必要があります。ここでは、分離戦略を採用し、コレスキー分解や球面分解に定式化が紹介されています。数値例は、3 変量の場合ですが以下では M 変量の場合で説明します</p>

<p>コレスキー分解 (Cholesky decomposition) による定式化</p>	<p>分散共分散行列$T_{M \times M}$ のコレスキー分解は次のように与えられます。</p> $T_{M \times M} = V^{1/2} R V^{1/2}$ <p>ここで、V は分散を要素とする対角行列、R は相関行列です。 さらに、R をコレスキー分解すると、L を上三角行列として、</p> $R_{M \times M} = L^T L$ <p>とパラメータ化できます。</p>
<p>球面分解 (spherical decomposition) による定式化</p>	<p>コレスキー分解で得られた L を球面分解により定式化することもできます。</p> $L_{l1} = \cos(\phi_{l2})$ $L_{l2} = \sin(\phi_{l2}) \cos(\phi_{l3})$ \vdots $L_{l,l-1} = \sin(\phi_{l2}) \sin(\phi_{l3}) \dots \cos(\phi_{ll})$ $L_{l,l} = \sin(\phi_{l2}) \sin(\phi_{l3}) \dots \sin(\phi_{ll})$ <p>例えば、$M=3$ のとき $L_{3 \times 3}$ は、次のようになります。</p> $L = \begin{bmatrix} 1 & \cos(\phi_{21}) & \cos(\phi_{31}) \\ 0 & \sin(\phi_{21}) & \sin(\phi_{31}) \cos(\phi_{32}) \\ 0 & 0 & \sin(\phi_{31}) \sin(\phi_{32}) \end{bmatrix}$ $R = L^T L = \begin{bmatrix} 1 & \cos(\phi_{21}) & \cos(\phi_{31}) \\ \cos(\phi_{21}) & 1 & * \\ \cos(\phi_{31}) & * & 1 \end{bmatrix}$ $* = \cos(\phi_{21}) \cos(\phi_{31}) + \sin(\phi_{21}) \sin(\phi_{31}) \cos(\phi_{32})$ <p>ϕ_{mn} に与える事前分布を $\text{Unif}(0, \pi)$ とすることで、R の非対角成分 (相関) の値域を $(-1, 1)$ に制約することができます</p>
<p>母数レベルのモデル②PNFによる定式化：構造をもたない分散共分散行列</p>	<p>試験間のモデル式 (12) を PNF で表します</p> $\begin{cases} \delta_{1i} \sim N(\eta_1, \psi_1^2) \\ \delta_{2i} \delta_{1i} \sim N(\eta_{2i}, \psi_2^2) \\ \eta_{2i} = \lambda_{20} + \lambda_{21} \delta_{1i} \\ \delta_{3i} \delta_{1i}, \delta_{2i} \sim N(\eta_{3i}, \psi_3^2) \\ \eta_{3i} = \lambda_{30} + \lambda_{31} \delta_{1i} + \lambda_{32} \delta_{2i} \end{cases}$
	<p>パラメータの関係は以下になります。導出の詳細は、C.7 を参照</p> $\begin{cases} \psi_1^2 = \tau_1^2 \\ \psi_2^2 = \tau_2^2 - \lambda_{21}^2 \tau_1^2 \\ \psi_3^2 = \tau_3^2 - \lambda_{31}^2 \tau_1^2 - \lambda_{32}^2 \tau_2^2 - 2\lambda_{31} \lambda_{32} \lambda_{21} \tau_1^2 \end{cases}, \quad \begin{cases} \lambda_{21} = \frac{\tau_2}{\tau_1} \rho^{12} \\ \lambda_{31} = \frac{\tau_3(\rho^{13} - \rho^{12} \rho^{23})}{\tau_1(1 - (\rho^{12})^2)} \\ \lambda_{32} = \frac{\tau_3(\rho^{23} - \rho^{12} \rho^{13})}{\tau_1(1 - (\rho^{12})^2)} \end{cases}$
<p>母数レベルのモデル③PNFによる定式化：構造を仮定した分散共分散行列</p>	<p>無構造のモデルは、多くのパラメータを推定する必要があり、推定のためにより大きなデータセットが必要になるかもしれません。パラメータの条件付き独立を仮定し、パラメータ数を減らすことを考えます。例えば、アウトカム 1 と 3 が条件付き独立であると仮定します。</p>

	$\begin{cases} \delta_{1i} \sim N(\eta_1, \psi_1^2) \\ \delta_{2i} \delta_{1i} \sim N(\eta_{2i}, \psi_2^2) \\ \eta_{2i} = \lambda_{20} + \lambda_{21} \delta_{1i} \\ \delta_{3i} \delta_{2i} \sim N(\eta_{3i}, \psi_3^2) \\ \eta_{3i} = \lambda_{30} + \lambda_{32} \delta_{2i} \end{cases}$
	<p>パラメータの関係は以下になります。導出の詳細は、C.7を参照</p> $\begin{cases} \psi_1^2 = \tau_1^2 \\ \psi_2^2 = \tau_2^2 - \lambda_{21}^2 \tau_1^2 \\ \psi_3^2 = \tau_3^2 - \lambda_{32}^2 \tau_2^2 \end{cases}, \begin{cases} \lambda_{21} = \frac{\tau_2}{\tau_1} \rho^{12} \\ \lambda_{32} = \frac{\tau_3}{\tau_2} \rho^{23} \end{cases}$ <p>アウトカム3と1への真の治療効果 δ_3, δ_1 の条件付き独立の仮定は、アウトカム2への治療効果 δ_2 で条件付けられ、それによって、共分散行列の逆行列（精度を表す）の $\{1, 3\}$ 要素はゼロになります。これは、δ_2 で条件付けた δ_2, δ_3 間の偏相関 $\rho^{13 2} = 0$ を意味します。</p> $\rho^{13 2} = \frac{\rho^{13} - \rho^{12} \rho^{23}}{\sqrt{1 - (\rho^{12})^2} \sqrt{1 - (\rho^{23})^2}} = 0 \text{ であり, これは, } \rho^{13} = \rho^{12} \rho^{23} \text{ を意味しま}$ <p>す。こうして、推定すべきモデルのパラメータ数を削減できただけでなく、試験間共分散行列の要素で表されるパラメータ間の関係も簡素化できました。</p>
TSDの数値例	<p>4.3 Example: rheumatoid arthritis</p> <p>HAQ, DAS-28, ACR を用いて分析する</p>
WinBUGS code	<p>C.2 WinBUGS code for multivariate meta-analysis with Cholesky decomposition of the between-studies covariance matrix with application to the example in rheumatoid arthritis</p> <p>C.3 WinBUGS code for multivariate meta-analysis with spherical decomposition of the between-studies covariance matrix with application to the example in rheumatoid arthritis</p> <p>C.4 WinBUGS code for TRMA with in PNF (structured model)</p>
事例	<p>4.5 Application to multiple surrogate endpoints</p> <p>血清 HIV-1 RNA と CD4+リンパ球を代替エンドポイントとした HIV 陽性の AIDS 患者の疾患進行の予測[101][102], 及び再発寛解型多発性硬化症患者の活動性病変 (active lesions) の数による障害進行 (disability progression) の予測[103]が紹介されています。</p> <p>また Bujkiewicz らは、複数の代替エンドポイントの評価のため、PNF による多変量メタアナリシスの手法を適用しました[72]。2変量メタアナリシスで行ったように、BRMA を3つ以上のアウトカムに拡張しました。Pozzi らも同様のアプローチを検討しています[104]。</p>

母数レベルのモデル②PNFによる定式化：構造をもたない分散共分散行列は、2022年に公表されたTSD 20第2版に基づく。

表 4 二変量変量効果 NMA モデル (bvNMA) : 対比データ

モデルの名称	二変量変量効果 NMA モデル (bvNMA) : 対比データ
TSD のセクション	5.1 Bivariate network meta-analysis (bvNMA) for contrast level data entry
試験レベルのモデル	<p>治療 k, b を比較した試験 i のアウトカム j を $Y_{(bk)ij}$ とします.</p> $\begin{pmatrix} Y_{(bk)i1} \\ Y_{(bk)i2} \end{pmatrix} \sim N \left(\begin{bmatrix} \delta_{(bk)i1} \\ \delta_{(bk)i2} \end{bmatrix}, \begin{bmatrix} \sigma_{(bk)i1}^2 & \sigma_{(bk)i1}\sigma_{(bk)i2}\rho_{(bk)wi} \\ \sigma_{(bk)i1}\sigma_{(bk)i2}\rho_{(bk)wi} & \sigma_{(bk)i2}^2 \end{bmatrix} \right)$ <p>治療効果 δ_i について, 移行性を仮定します.</p> $\begin{pmatrix} \delta_{(kl)i1} \\ \delta_{(kl)i2} \end{pmatrix} = \begin{pmatrix} \delta_{(bl)i1} - \delta_{(bk)i1} \\ \delta_{(bl)i2} - \delta_{(bk)i2} \end{pmatrix}$
母数レベルのモデル	$\begin{pmatrix} \delta_{(bk)i1} \\ \delta_{(bk)i2} \end{pmatrix} \sim N \left(\begin{bmatrix} d_{(bk)1} \\ d_{(bk)2} \end{bmatrix}, \begin{bmatrix} \tau_{(bk)1}^2 & \tau_{(bk)1}\tau_{(bk)2}\rho_{(bk)} \\ \tau_{(bk)1}\tau_{(bk)2}\rho_{(bk)} & \tau_{(kl)2}^2 \end{bmatrix} \right)$ <p>δ_i についての移行性から, 一次の一致性が導かれます.</p> $\begin{pmatrix} d_{(kl)i1} \\ d_{(kl)i2} \end{pmatrix} = \begin{pmatrix} d_{(bl)1} - d_{(bk)1} \\ d_{(bl)2} - d_{(bk)2} \end{pmatrix}$
TSD の数値例	<p>5.4 Example: relapsing remitting multiple sclerosis (RRMS)</p> <p>Melendez-Torres らは, 成人の再発寛解型多発性硬化症患者を対象とし, ベータインターフェロンとグラチラマー酢酸塩を治療選択肢として, システマティックレビュー及びメタアナリシスを行いました[105]. システマティックレビューで特定された 32 試験から, 複数のアウトカムが抽出されました: 年間再発率, 障害の進行, 追跡終了時まで再発のなかった患者の割合, 治療関連の有害事象, ランダム化後 24 か月までの有害事象による治療中止の割合. ここでは, 2 変量のネットワークメタアナリシスを説明するために, 再発なしの割合と有害事象による治療中止の割合を使用します. いずれかのアウトカムが報告された試験は 32 試験中 22 試験, 再発なしの割合は 22 試験で 8 つの治療比較, 両アウトカムとも報告されたのは 11 試験で 6 つの治療比較でした.</p> <p>TSD 20 では, 多変量正規分布に基づくモデルを扱うため, 連続変数で近似的に正規分布に従うアウトカム変数が必要です. RRMS の数値例では, 二値変数データを治療比較のオッズ比に変換します (ここで, 応答が 0 の群には 0.5 を加えました).</p> <p>試験内相関のデータは得られていないため, 試験内相関に複数のモデルをあてはめます.</p> <p>・モデル 1, 試験内相関及び試験間相関をゼロ (定数) に設定し, アウトカムが無相関と仮定します. このモデルは, 個別の単変量ネットワークメタアナリシスのモデルをあてはめることと同等です.</p>

	<p>・モデル2, 試験内相関をゼロ (定数) に設定し, アウトカムが試験内では無相関と仮定します. 試験間相関は, 事前分布を一様分布 Unif (-1, 1) として推定します (モデル3, 4も同様).</p> <p>・モデル3, 試験内相関を正と仮定し, 事前分布を一様分布 Unif (0, 1) として推定します. より効果的な治療 (再発の少ない治療) は, より有害事象が多く, よって有害事象による中止もより多い臨床的な仮定を反映しています.</p> <p>・モデル4, 試験内相関を一様分布 Unif (-1, 1) として, 全く情報が無いことを反映しています.</p> <p>これらは説明のためのモデルとして設定したものです. 他にも試験内相関を負と仮定して, Unif (-1, 0) とすることも考えられます. これは, 治療効果が小さいと有害事象による治療中止も少ないという想定を反映します.</p>
WinBUGS code	D.2 WinBUGS code for multivariate NMA applied to the example in multiple sclerosis

表 5 二変量変量効果 NMA モデル (bvNMA) : 群別データ

モデルの名称	二変量変量効果 NMA モデル (bvNMA) : 群別データ
TSD のセクション	5.2 Bivariate network meta-analysis (bvNMA) for arm level data entry
試験レベルのモデル	<p>試験 i の治療 k のアウトカム j を Y_{kij} とします</p> $\begin{pmatrix} Y_{ki1} \\ Y_{ki2} \end{pmatrix} \sim N \left(\begin{bmatrix} \theta_{ki1} \\ \theta_{ki2} \end{bmatrix}, \begin{bmatrix} \sigma_{ki1}^2 & \sigma_{ki1}\sigma_{ki2}\rho_{kwi} \\ \sigma_{ki1}\sigma_{ki2}\rho_{kwi} & \sigma_{ki2}^2 \end{bmatrix} \right)$ <p>推移律を仮定し, 試験 i の対照治療 b のアウトカム j の期待値を μ_{bij} とし, 治療 k の対照治療 b に対する治療効果を $\delta_{(bk)i1}$ とすると, 治療 k のアウトカムの期待値 θ_{kij} は,</p> $\begin{pmatrix} \theta_{ki1} \\ \theta_{ki2} \end{pmatrix} = \begin{pmatrix} \mu_{bi1} + \delta_{(bk)i1} \\ \mu_{bi2} + \delta_{(bk)i2} \end{pmatrix}, \text{ ただし, } k=b \text{ のとき, } \delta_{(bk)ij} = 0$
母数レベルのモデル	$\begin{pmatrix} \delta_{(bk)i1} \\ \delta_{(bk)i2} \end{pmatrix} \sim N \left(\begin{bmatrix} d_{(bk)1} \\ d_{(bk)2} \end{bmatrix}, \begin{bmatrix} \tau_{(bk)1}^2 & \tau_{(bk)1}\tau_{(bk)2}\rho_{bk} \\ \tau_{(bk)1}\tau_{(bk)2}\rho_{bk} & \tau_{(bk)2}^2 \end{bmatrix} \right)$ <p>全体のリファレンス治療を 1 とすると, 次の関係式が成り立ちます</p> $d_{(bk)j} = d_{(1k)j} - d_{(1b)j}$
母数レベルのモデル : 試験間の分散共分散行列に交換可能性を仮定	<p>しばしば, 試験間の分散共分散行列に均一性 (homogeneity) を仮定し, パラメータ数を減らしたモデルが用いられることがあります</p> $\begin{pmatrix} \delta_{(bk)i1} \\ \delta_{(bk)i2} \end{pmatrix} \sim N \left(\begin{bmatrix} d_{(bk)1} \\ d_{(bk)2} \end{bmatrix}, \begin{bmatrix} \tau_1^2 & \tau_1\tau_2\rho \\ \tau_1\tau_2\rho & \tau_2^2 \end{bmatrix} \right)$
WinBUGS code	なし

表 6 多変量変量効果 NMA モデル (mvNMA) : 群別データ

モデルの名称	多変量変量効果 NMA モデル (mvNMA) : 群別データ
TSD のセクション	5.3 Multivariate network meta-analysis (mvNMA) for arm level data entry
試験レベルのモデル	<p>試験 i, 治療 k, アウトカム $1 \sim M$ について,</p> $\begin{pmatrix} Y_{ki1} \\ \vdots \\ Y_{kiM} \end{pmatrix} \sim N \left(\begin{bmatrix} \theta_{ki1} \\ \vdots \\ \theta_{kiM} \end{bmatrix}, \begin{bmatrix} \sigma_{ki1}^2 & \dots & \sigma_{ki1}\sigma_{kiM}\rho_{ki}^{1M} \\ \vdots & \ddots & \vdots \\ \sigma_{ki1}\sigma_{kiM}\rho_{ki}^{1M} & \dots & \sigma_{kiM}^2 \end{bmatrix} \right)$ <p>通常, 分散共分散行列は IPD から推定した上で, 既知とされます.</p>
母数レベルのモデル	<p>bvNMA モデルと同様に,</p> $\begin{pmatrix} \theta_{ki1} \\ \vdots \\ \theta_{kiM} \end{pmatrix} = \begin{pmatrix} \mu_{bi1} + \delta_{(bk)i1} \\ \vdots \\ \mu_{biM} + \delta_{(bk)iM} \end{pmatrix}, \text{ ただし, } k=b \text{ のとき, } \delta_{(bk)ij} = 0$
	<p>試験間の分散共分散行列に均一性を仮定した場合,</p> $\begin{pmatrix} \delta_{(bk)i1} \\ \vdots \\ \delta_{(bk)iM} \end{pmatrix} \sim N \left(\begin{bmatrix} d_{(bk)1} \\ \vdots \\ d_{(bk)M} \end{bmatrix}, \begin{bmatrix} \tau_1^2 & \dots & \tau_1\tau_M\rho \\ \vdots & \ddots & \vdots \\ \tau_1\tau_M\rho & \dots & \tau_M^2 \end{bmatrix} \right)$
	多群試験への拡張は, Appendix D.1 を参照してください.
WinBUGS code	なし

アウトカムが1変量NMAでは、(通常のメタアナリシスを利用して、各対比のデータを個別に解析する場合と比較して)他の治療の対比から情報を借りることができます。したがって、多変量メタアナリシスを利用して、多変量ネットワークメタアナリシスに比して、情報を借りることによる追加的な強度は期待できないかもしれません。しかし、1つのアウトカムに関するネットワークが疎である場合、他のアウトカムから情報を借りることで、推定値の不確実性を減らすことができます。さらに、医療経済意思決定モデルで使用される相関のある平均治療効果の同時事後分布が、費用効果モデルのより適切な推定値をもたらすかもしれません[106]。2変量NMAのもう1つの新しい応用例は、4.8節で議論したように、Bujkiewiczら[43]が開発した代替エンドポイント評価の分野になります。このような方法論は、関心のある1つまたは複数の介入について、意志決定に利用されるエンドポイントに対する効果がどの研究からも得られないが、代替エンドポイントに関するエビデンスは利用可能な場合に有用です。その場合、特定の介入について関心のあるアウトカムの治療効果を予測することができます。多変量NMAの他の手法も報告されています。また、Hongら[48]は、対照群との相対的な治療効果と、単群での絶対的な治療効果の両方について、個々の患者データを用いた複数アウトカムのNMAを提案しました。また、Hongら[107]は多変量NMAのベイズのフレームワークも構築しています。これらの多変量NMAモデルは、ネットワークにおける治療効果の一致性の仮定に基づいており、Lu and Ades[108]により導入されたアプローチを拡張したものです。また、Jacksonら[109]は、一致性が成立しない場合を考慮した多変量NMAモデルを開発しました。治療効果のネットワークにおける一致性の仮定が満たされず、一致しない効果を説明する追加の項をモデルに含めました。言い換えれば、研究間の異質性を考慮しても、異なる形式の直接エビデンスと間接エビデンスが一致しないことがあります。一致性がないこと(inconsistency)を試験デザインと治療効果の交互作用項を用いて考慮します。ここで、試験デザインとは、AとBの比較試験や、治療法C、D、Eを比較する3群試験など、試験の構成要素を指します。これらの方法の目的は、ネットワークにおける相関のある複数のアウトカム及び治療法の対比に関する統合した治療効果を得ることです。

6. 議論と拡張

6.1 変量効果の正規性の仮定

前節までに解説してきた方法は、モデル化された真の治療効果が研究間で異なる（しかし類似している）という仮定を反映する変量効果モデルでした。真の治療効果の違いは、母集団が異なること、治療法が試験間で異なること、あるいはアウトカム定義又は追跡期間における異質性に起因する可能性があります[44]。一般的に、治療効果の類似性を反映するため、試験間の変量効果に正規分布が仮定されます。しかし、両方のアウトカムに対する真の治療効果（2つのアウトカムの対数オッズ比など）が正規分布であるという仮定は、必ずしも妥当とはいえない場合があります[110]。モデル化されたデータの正規性からの逸脱を扱う場合、この仮定は、モデル化と推論を制限してしまう可能性があります[111]。特に、新しい試験又は新しい集団における治療効果の潜在的な大きさについて予測的推論を行う場合です。例えば、Marshall and Spiegelhalter が議論したように、変量効果に関する正規性の仮定が不適切な場合、真の治療効果の「過剰縮小」につながる可能性があります[112]。

正規性の仮定を緩和する1つの方法は、例えば、Smith, Spiegelhalter and Thomas[113] [106]、又は Lee and Thompson[111]が推奨する t 分布を使用することです。正規分布とは対照的に、 t 分布はその裾により重みがあるため、外れ値のような極端な治療効果をモデル化するのに適しています[112]。データの分布が、例えば、二峰性または歪んでいる場合、正規分布の畳込み（convolution of normal distributions）[114]、又は Lee and Thompson[111] が提案したような歪んだ t 分布など、他のアプローチを検討することができます。

6.2 二値変数、ネストされた、又は相互に排他的なアウトカム

多くの場合、データは複数の二値変数から収集されます。標準的なアプローチは、正規近似を用いて対数オッズ比尺度（または単群データの対数オッズ尺度）でモデル化することです。例えば、関節リウマチにおける ACR20 の例（TSD 20.2.3 節）、又は多発性硬化症における多変量ネットワークメタアナリシスの例（TSD 20 付録 D.4）を参照してください。二項尤度を用いた厳密な方法は、特に、メタアナリシスで試験数が少ない場合、又はイベントの確率が 0 または 1 に近い場合に、望ましいかもしれません。しかし、二変数の場合、試験内相関を無視したモデルに限定されるため、偏った結果になる可能性があります[51]。この問題は、Chen ら[115]がさらに議論し、相関のある二項データをモデル化する代替手法を提案しています。しかし、二項モデルは、アウトカムが相互に排他的であるか、又は「部分集合」の関係がある場合（ネストされた場合）に利用することができ、Trikalinos ら[116]は、カテゴリ

データの多変量メタアナリシスモデルについても論じています。

6.3 ベイズの方法と頻度論の方法, 及び関連ソフトウェア

TSD 20 では, ベイズの方法を利用して, 複数のアウトカムをモデル化しました。ベイズの方法は, 複数パラメータの不確実性を柔軟にモデル化するのに最も適しています。多変量メタアナリシスモデルは, 頻度論の方法を使用して実装することもできます。例えば, Stata では, コマンド `mvmeta`[117]を使用して実装できます。Bujkiewicz ら[55]は, `mvmeta` コマンドと一緒に使用できる代替エンドポイント評価のクロスバリデーションの手順のための追加の Stata コードを提供しています。ただし, 標準形式の BRMA のみが Stata で使用できます (BRMA PNF 又は代替エンドポイントの Daniels and Hughes によるモデルは利用できません)。多変量メタアナリシスは, `mvmeta` パッケージを利用して R で実行することもできます[118]。多変量ネットワークメタアナリシスのためのソフトウェアは, Jackson ら[109]により開発された R で利用可能です。一致性が成立していない場合を考慮したモデルを利用できますが, 一致性を仮定したモデルの推定も可能です。Polanin ら[119]は, R における (多変量メタアナリシスも含む)メタアナリシスについて, 他のパッケージに対するレビューを提供しています。

6.4 その他の応用分野(QOL 値など)

多変量メタアナリシスは幅広い領域で応用できます。治療効果データに加えて, 有害事象データを含めることもできます (例えば, 多発性硬化症における多変量ネットワークメタアナリシス例, TSD 20 5.4 節参照)。同様に, HRQoL に対応する一般的な尺度 (又は包括的な尺度) (例えば, EQ-5D) の推定値を得るために, 2つを同時にモデル化 (joint model) するか, マッピング技術により, 多変量メタアナリシスを利用することができます。Bujkiewicz ら[56]は, 多変量メタアナリシスの利用により, 疾患特異的 HRQoL 尺度がより正確に推定され, それを EQ-5D などの一般的な尺度へのマッピングする場合において, EQ-5D などの一般的な尺度の推定値も, より正確に得られることを示しました。それは費用対効果モデルに基づいた意思決定の不確実性を低減する (あるいは, 英国では新しい医療技術の償還に関する決定が変わる) 可能性を示しています。多変量メタアナリシスは, 臨床効果と HRQoL アウトカム (疾患特異的な尺度と EQ-5D のような一般的な尺度) を同時にモデル化するにも利用でき, 全てのアウトカムに対して統合した推定値が得られるだけでなく, 代替エンドポイントの治療効果から最終的な臨床アウトカムの治療効果を予測したときと同様の方法で, 価値のある尺度に関する推定値を予測することができます。ある種の多変量メタアナリシスを含むその他のマッピング手法が Lu ら[120]により提案されています。

Dias ら[121]は、意思決定における多変量メタアナリシスの他の多くの応用について論じています。アウトカム間の機能的関係（functional relationship）を捉えることにより、エビデンス構造をモデル化することができる多くのシナリオを説明しています。例えば、疾患の自然史（disease natural history）をモデル化する場合における、アウトカム間の関係を導く「エビデンスの連鎖」構造、及び異なる時点において異なるネットワーク構造を生成する複数時点での利用が含まれます。

7. まとめ

TSD 20 の推奨事項

- 関心のあるアウトカムについては、scoping stage（補足：本邦の制度における分析前協議に相当）の段階で慎重に検討する必要があります
- 複数のエンドポイントに対する治療効果の相関は、頑健な治療効果の推定値のために、同時に分析されるべきです
- 試験数が少ない場合や、アウトカムデータの欠測がある場合は、精度がより向上する可能性があります、多変量メタアナリシスの使用を推奨します
- 相関のあるアウトカム同士の多変量事後分布は、経済モデルへの入力として適切かつ有用と考えます
- 代替エンドポイントは、多変量メタアナリシスの手法を用いて検証されるべきです

ベイズ流の多変量メタアナリシスのモデルは試験レベルの治療効果の交換可能性や治療効果レベルで試験間の分散共分散行列の交換可能性を仮定するので、それらの仮定が妥当かどうか検討する必要があります。また、真のエンドポイントが得られていない状況で、代替エンドポイントと他の治療のエビデンスから多変量メタアナリシスによって真のエンドポイントでの治療効果を推定する際には、代替エンドポイントと真のエンドポイントの患者レベルの相関のみでは不十分で、代替エンドポイントに対する治療効果と真のエンドポイントに対する治療効果が関連し、定量的に推定できるかどうかについても注意が必要です[122][95][9]。

日本の費用対効果評価ガイドライン第3版[14]には、代替エンドポイントについての言及はありませんが、HTAでの代替エンドポイントの利用に関して具体的なガイダンスを提供している海外のHTA機関も存在します[123][82]。HTAにおける代替エンドポイントの妥当性評価や受け入れ可能性の評価については、本報告書の範囲を超えますが、必要に応じて文献等をご参照ください。

多変量メタアナリシスを適切に用いることで、より多くの情報をエビデンスの統合に利用できる可能性があり、また、代替エンドポイントでの効果の検証によって承認された薬剤の真のエンドポイントの効果を定量的に推定することができれば、医療経済評価に有用な情報を提供することができる可能性があります。そうすることによってヘルスケアに関するより良い意思決定につながる可能性があります。本報告書が少しでもみなさまのお役に立てば幸いです。最後までお読み頂き、ありがとうございました。

8. 引用文献

1. 宇沢弘文. 社会的共通資本. 2000 岩波新書.
2. 高橋伸彰. 経済学の貧困と経済の危機に関する考察— 宇沢弘文の思想に学ぶ —. 立命館国際関係研究 28(4) 61-77 2016年3月.
3. 医薬品、医療機器及び再生医療等製品の費用対効果評価に関する取扱いについて. (2022).
Available: https://c2h.niph.go.jp/tools/system/notification_cost-effectiveness.pdf.
4. National Institute for Health and Care Excellence. Available: <https://www.nice.org.uk/>.
5. NICE Decision Support Unit. Available: <https://nicedsu.org.uk/>.
6. 日本製薬工業協会. NICE TSD Evidence Synthesis シリーズの紹介： ネットワークメタアナリシスを中心として. 2020. Available:
https://www.jpma.or.jp/information/evaluation/results/allotment/nice_tsd_evidence_synthesis.html
7. Bujkiewicz, S., Achana, F., Papanikos, T., Riley, R.D., Abrams, K.R. NICE DSU Technical Support Document 20: Multivariate meta-analysis of summary data for combining treatment effects on correlated outcomes and evaluating surrogate endpoints. 2019; available from <http://www.nicedsu.org.uk>
(注：本報告書作成時点では上記URLで公開されている文書はVer2.0である)
8. Ciani O, Grigore B, Taylor RS. Development of a framework and decision tool for the evaluation of health technologies based on surrogate endpoint evidence. *Heal Econ (United Kingdom)*. 2022;31(S1):44–72.
9. Ciani O, Grigore B, Blommestein H, de Groot S, Möllenkamp M, Rabbe S, et al. Validity of Surrogate Endpoints and Their Impact on Coverage Recommendations: A Retrospective Analysis across International Health Technology Assessment Agencies. *Med Decis Mak*. 2021;41(4):439–52.
10. Drummond M, Tarricone R, Torbica A. European union regulation of health technology assessment: what is required for it to succeed? *Eur J Heal Econ [Internet]*.

- 2022;23(6):913–5. Available from: <https://doi.org/10.1007/s10198-022-01458-6>
11. Lux MP, Ciani O, Dunlop WCN, Ferris A, Friedlander M. The impasse on overall survival in oncology reimbursement decision-making: How can we resolve this? *Cancer Manag Res.* 2021;13(November):8457–71.
 12. Mitra-Majumdar M, Gunter SJ, Kesselheim AS, Brown BL, Joyce KW, Ross M, et al. Analysis of Supportive Evidence for US Food and Drug Administration Approvals of Novel Drugs in 2020. *JAMA Netw Open.* 2022;5(5):1–11.
 13. Table of Surrogate Endpoints That Were the Basis of Drug Approval or Licensure Content current as of: 02/28/2022 Available: <https://www.fda.gov/drugs/development-resources/table-surrogate-endpoints-were-basis-drug-approval-or-licensure>.
 14. 国立保健医療科学院保健医療経済評価研究センター. 中央社会保険医療協議会における費用対効果評価の分析ガイドライン 第3版. 2022; Available from: https://c2h.niph.go.jp/tools/guideline/guideline_ja.pdf
 15. Dias, S., Welton, N.J., Sutton, A.J. & Ades, A.E. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials. 2011; last updated September 2016; available from <http://www.nicedsu.org.uk>
 16. Dias, S., Sutton, A.J., Welton, N.J., Ades, A.E. NICE DSU Technical Support Document 3: Heterogeneity: subgroups, meta-regression, bias and bias-adjustment. 2011; last updated April 2012; available from <http://www.nicedsu.org.uk>
 17. Dias, S., Welton, N.J., Sutton, A.J., Caldwell, D.M., Lu, G. & Ades, A.E. NICE DSU Technical Support Document 4: Inconsistency in Networks of Evidence Based on Randomised Controlled Trials. 2011; last updated April 2014; available from <http://www.nicedsu.org.uk>
 18. Dias, S., Welton, N.J., Sutton, A.J., Ades, A.E. NICE DSU Technical Support Document 5: Evidence synthesis in the baseline natural history model. 2011; last updated April 2012; available from <http://www.nicedsu.org.uk>
 19. 保健医療経済評価研究センター. Available: <https://c2h.niph.go.jp/>.
 20. 鎌江伊三夫. (2016). 医療技術評価ワークブック. じほう.
 21. 五十嵐中, 佐條麻里. (2014). 「薬剤経済」わかりません!! . 東京図書.
 22. Drummond, M., Sculpher, M. J., Claxton, K., Stoddart, G. L., Torrance, G.W., (監訳) 久繁哲徳, 橋本英樹. (2017). 保健医療の経済評価 第 4 版. 篠原出版新社.
 23. 日本製薬工業協会. 医薬品の価値の科学的な評価 – データサイエンス担当者のための費用対効果評価の現状と手法の解説 – . 2016; Available: http://www.jpma.or.jp/medicine/shinyaku/tiken/allotment/2014ds_tf3.html
 24. 日本製薬工業協会. 費用対効果評価の実践 – NICE 技術評価ガイダンスの事例から分析モデルのバリエーションを学ぶ – . 2018;

Available: <https://www.jpma.or.jp/information/evaluation/results/allotment/cost-effectiveness.html>

25. 日本製薬工業協会. 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 –. 2019;
Available:
https://www.jpma.or.jp/information/evaluation/results/allotment/network_meta_analysis.html
26. Riley RD, Jackson D, Salanti G, Burke DL, Price M, Kirkham J, et al. Multivariate and network meta-analysis of multiple outcomes and multiple treatments : rationale , concepts , and examples. *BMJ* 2017;358:j3932
27. Jones AP, Riley RD, Williamson PR, Whitehead A. Meta-analysis of individual patient data versus aggregate data from longitudinal clinical trials. *Clin Trials*. 2009;6(1):16–27. Available from: <https://pubmed.ncbi.nlm.nih.gov/19254930/>
28. Thompson JR, Minelli C, Abrams KR, Tobin MD, Riley RD. Meta-analysis of genetic studies using Mendelian randomization - A multivariate approach. *Stat Med*. 2005;24(14):2241–54.
29. The Fibrinogen Studies Collaboration. Systematically missing confounders in individual participant data meta-analysis of observational cohort studies. *Stat Med*. 2009;28:1218–37.
30. Snell KIE, Hua H, Debray TPA, Ensor J, Look MP, Moons KGM, et al. Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model. *J Clin Epidemiol* [Internet]. 2016;69:40–50. Available from: <http://dx.doi.org/10.1016/j.jclinepi.2015.05.009>
31. Riley RD, Elia EG, Malin G, Hemming K, Price MP. Multivariate meta-analysis of prognostic factor studies with multiple cut-points and/or methods of measurement. *Stat Med*. 2015;34(17):2481–96.
32. Kirkham JJ, Riley RD, Williamson PR. A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. *Stat Med*. 2012;31(20):2179–95.
33. Jackson D, Riley R, White IR. Multivariate meta-analysis: Potential and promise. *Stat Med*. 2011;30(20):2481–98.
34. Van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: Multivariate approach and meta-regression. *Stat Med*. 2002;21(4):589–624.
35. Nam IS, Mengersen K, Garthwaite P. Multivariate meta-analysis. *Stat Med*. 2003;22(14):2309–33.
36. Riley RD, Abrams KR, Lambert PC, Sutton AJ, Thompson JR. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Stat Med*. 2007;26(1):78–97.
37. Bujkiewicz S, Thompson JR, Sutton AJ, Cooper NJ, Harrison MJ, Symmons DPM, et al. Multivariate meta-analysis of mixed outcomes: A Bayesian approach. *Stat Med*. 2013;32(22):3926–43.
38. Wei Y, Higgins JPT. Bayesian multivariate meta-analysis with multiple outcomes. *Stat*

- Med. 2013;32(17):2911–34.
39. Jackson D, White IR, Price M, Copas J, Riley RD. Borrowing of strength and study weights in multivariate and network meta-analysis. *Stat Methods Med Res.* 2017;26(6):2853–68.
 40. Spiegelhalter DJ, Abrams KR, Myles. JP. Bayesian approaches to clinical trials and health-care evaluation, volume 13. John Wiley & Sons, 2004. 2004.
 41. Spiegelhalter DJ, Best NG. Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Stat Med.* 2003;22(23):3687–709.
 42. Ashby D, Smith AFM. Evidence-based medicine as Bayesian decision-making. *Stat Med.* 2000;19(23):3291–305.
 43. Bujkiewicz S, Jackson D, Thompson JR, Turner RM, Städler N, Abrams KR, et al. Bivariate network meta-analysis for surrogate endpoint evaluation. *Stat Med.* 2019;38(18):3322–41.
 44. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2009;172(1):137–59.
 45. Efthimiou O, Mavridis D, Cipriani A, Leucht S, Bago P, Salanti G. An approach for modelling multiple correlated outcomes in a network of interventions using odds ratios. *Stat Med.* 2014;33(13):2275–87.
 46. Achana FA, Cooper NJ, Bujkiewicz S, Hubbard SJ, Kendrick D, Jones DR, et al. Network meta-analysis of multiple outcome measures accounting for borrowing of information across outcomes. *BMC Med Res Methodol.* 2014;14(1):1–16.
 47. Efthimiou O, Mavridis D, Riley RD, Cipriani A, Salanti G. Joint synthesis of multiple correlated outcomes in networks of interventions. *Biostatistics.* 2015;16(1):84–97.
 48. Hong H, Fu H, Price KL, Carlin BP. Incorporation of individual-patient data in network meta-analysis for multiple continuous endpoints, with application to diabetes treatment. *Stat Med.* 2015;34(20):2794–819.
 49. Hong H, Chu H, Zhang J, Carlin BP. A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons. *Res Synth Methods* ;7(1):6–22. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1002/jrsm.1153>
 50. Jackson D, Bujkiewicz S, Law M, Riley RD, White IR. A matrix-based method of moments for fitting multivariate network meta-analysis models with multiple outcomes and random inconsistency effects. *Biometrics.* 2018;74(2):548–56.
 51. Riley RD. Multivariate meta-analysis: the effect of ignoring within-study correlation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 172(4):789–811.
 52. Daniels MJ, Hughes MD. Meta-analysis for the evaluation of potential surrogate markers. *Stat Med.* 1997;16(17):1965–82.
 53. Wei Y, Higgins JP. Estimating within-study covariances in multivariate meta-analysis with multiple outcomes. *Stat Med.* 2013;32(7):1191–205.

54. Riley RD, Thompson JR, Abrams KR. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics*. 2008;9(1):172–86.
55. Bujkiewicz S, Thompson JR, Spata E, Abrams KR. Uncertainty in the Bayesian meta-Analysis of normally distributed surrogate endpoints. *Stat Methods Med Res*. 2017;26(5):2287–318.
56. Bujkiewicz S, Thompson JR, Sutton AJ, Cooper NJ, Harrison MJ, Symmons DPM, et al. Use of bayesian multivariate meta-analysis to estimate the HAQ for mapping onto the EQ-5D questionnaire in rheumatoid arthritis. *Value in Health*. 2014;17(1):109–15.
57. Copas JB, Jackson D, White IR, Riley RD. The role of secondary outcomes in multivariate meta-analysis. *J R Stat Soc Ser C Appl Stat*. 2018;67(5):1177–205.
58. Efthimiou O, Debray TPA, van Valkenhoef G, Trelle S, Panayidou K, Moons KGM, et al. GetReal in network meta-analysis: a review of the methodology. *Res Synth Methods*. 2016;7(3):236–63.
59. Riley RD, Ensor J, Jackson D, Burke DL. Deriving percentage study weights in multi-parameter meta-analysis models: with application to meta-regression, network meta-analysis and one-stage individual participant data models. *Stat Methods Med Res*. 2018;27(10):2885–905.
60. Seaman S, Galati J, Jackson D, Carlin J. What is meant by “missing at random”? *Stat Sci*. 2013;28(2):257–68.
61. Nixon RM, Duffy SW, Fender GRK. Imputation of a true endpoint from a surrogate: Application to a cluster randomized controlled trial with partial information on the true endpoint. *BMC Med Res Methodol*. 2003;3:1–11.
62. D’Agostino J. Debate: The slippery slope of surrogate outcomes. *Curr Control Trials Cardiovasc Med*. 2000;1(2):76–8.
63. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ*. 2010;340(7747):637–40.
64. Dwan K, Altman DG, Clarke M, Gamble C, Higgins JPT, Sterne JAC, et al. Evidence for the Selective Reporting of Analyses and Discrepancies in Clinical Trials: A Systematic Review of Cohort Studies of Clinical Trials. *PLoS Med*. 2014;11(6):1–22.
65. National Institute for Health and Care Excellence. Continuous positive airway pressure for the treatment of obstructive sleep apnoea/hypopnoea syndrome
Technology appraisal guidance [TA139] Published: 26 March 2008 Last updated: 20 August 2021
<https://www.nice.org.uk/guidance/ta139>
66. National Institute for Health and Care Excellence. TNF-alpha inhibitors for ankylosing spondylitis and non-radiographic axial spondyloarthritis
Technology appraisal guidance [TA383] Published: 01 February 2016
<https://www.nice.org.uk/guidance/ta383>
67. Tanaka S, Oba K, Yoshimura K, Teramukai S. Statistical Criteria for Surrogate Endpoint and Applications: A Review. *Japanese J Biometrics*. 2010;31(1):23–48.

68. Molenberghs G, Buyse M, Burzykowski T. The evaluation of surrogate endpoints. Springer. 2006.
69. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics*. 2000;1(1):49–67.
70. Burzykowski T, et al. Validation of Surrogate End Points in Multiple Randomized Clinical Trials with Failure Time End Points. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2001;50(4):405–22.
71. Renfro LA, Shi Q, Sargent DJ, Carlin BP. Bayesian adjusted r^2 for the meta-analytic evaluation of surrogate time-to-event endpoints in clinical trials. *Stat Med*. 2012;31(8):743–61.
72. Bujkiewicz S, Thompson JR, Riley RD, Abrams KR. Bayesian meta-analytical methods to incorporate multiple surrogate endpoints in drug development process. *Stat Med*. 2016;35(7):1063–89.
73. Taylor RS, Elston J. The use of surrogate outcomes in model-based cost-effectiveness analyses: a survey of UK Health Technology Assessment reports. *Health Technol Assess*. 2009 Jan 9;13(8):1–72. Available from: <https://www.journalslibrary.nihr.ac.uk/hta/hta13080/>
74. Ciani O, Buyse M, Drummond M, Rasi G, Saad ED, Taylor RS. Time to Review the Role of Surrogate End Points in Health Policy: State of the Art and the Way Forward. *Value Heal*. 2017;20(3):487–95. Available from: <http://dx.doi.org/10.1016/j.jval.2016.10.011>
75. Tan SH, Abrams KR, Bujkiewicz S. Bayesian Multiparameter Evidence Synthesis to Inform Decision Making: A Case Study in Metastatic Hormone-Refractory Prostate Cancer. *Med Decis Mak*. 2018;38(7):834–48.
76. European Medicines Agency. Final report from the emea/chmp-think-tank group on innovative drug development. available at <http://bit.ly/2rCmLAH>, 2007.
77. Coons SJ. The FDA’s critical path initiative: A brief introduction. *Clin Ther*. 2009 Nov 1;31(11):2572–3.
78. Kim C, Prasad V. Cancer drugs approved on the basis of a surrogate end point and subsequent overall survival: An analysis of 5 years of us food and drug administration approvals. *JAMA Intern Med*. 2015;175(12):1992–4.
79. European Medicines Agency. European public assessment report of lartruvo (olaratumab). available at <https://bit.ly/2VnLv9Z>, 2016. 66
80. FDA Center for Drug Evaluation and Research. Regulatory Decision to Withdraw Avastin (bevacizumab) First- line Metastatic Breast Cancer Indication. available at <https://bit.ly/2HpeLrR>, 2010.
81. National Institute for Health and Care Excellence. Venetoclax in combination with rituximab for treating relapsed or refractory chronic lymphocytic leukaemia [id1097]. available at <https://www.nice.org.uk/guidance/ta561>, 2019.
82. EUnetHTA. Endpoints used in Relative Effectiveness Assessment: Surrogate Endpoints. Adapted version (2015) <https://www.eunetha.eu/wp->

content/uploads/2018/03/surrogate_endpoints.pdf

83. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. available at <https://www.nice.org.uk/process/pmg9/chapter/foreword>, 2013.
84. Sormani M, Bonzano L, Roccatagliata L, Mancardi G, Uccelli A, Bruzzi P. Surrogate endpoints for edss worsening in multiple sclerosis a meta-analytic approach. *Neurology*. 2010;75(4):302-9.
85. International Conference on Harmonisation guidelines for the conduct of clinical trials for the registration of drugs. E9: statistical principles for clinical trials, Published 1998. Accessed February 22, 2019.
86. Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. Users' guides to the medical literature: XIX. Applying clinical trial results A. How to use an article measuring the effect of an intervention on surrogate end points. *Jama*. 1999;282(8):771–8.
87. Lassere MN, Johnson KR, Boers M, Tugwell P, Brooks P, Simon L, et al. Definitions and validation criteria for biomarkers and surrogate endpoints: Development and testing of a quantitative hierarchical levels of evidence schema. *J Rheumatol*. 2007;34(3):607–15.
88. Fleming TR, DeMets DL. Surrogate End Points in Clinical Trials: Are We Being Misled? *Ann Intern Med*. 1996 Oct 1;125(7):605–13.
89. Joffe MM, Greene T. Related causal frameworks for surrogate outcomes. *Biometrics*. 2009;65(2):530.
90. VanderWeele TJ. Surrogate Measures and Consistent Surrogates. *Biometrics*. 2013;69(3):569–73.
91. Buyse M, Molenberghs G, Paoletti X, Oba K, Alonso A, Van der Elst W, et al. Statistical evaluation of surrogate endpoints with examples from cancer clinical trials. *Biometrical J*. 2016 ;58(1):104–32.
92. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc Ser B Stat Methodol*. 2002;64(4):583–616.
93. Lassere MN, Johnson KR, Schiff M, Rees D. Is blood pressure reduction a valid surrogate endpoint for stroke prevention? an analysis incorporating a systematic review of randomised controlled trials, a by-trial weighted errors-in-variables regression, the surrogate threshold effect (STE) and the b. *BMC Med Res Methodol*. 2012;12:27.
94. Institute for Quality and Efficiency in Health Care. Validity of surrogate endpoints in oncology. executive summary of rapid report a10-05, version 1.1., <https://www.ncbi.nlm.nih.gov/books/nbk198799/>, Published 2011. Accessed December 7, 2018.
95. Alonso A, Bigirimurame T, Burzykowski T, Buyse M, Molenberghs G, Muchene L, et al. *Applied Surrogate Endpoint Evaluation Methods with SAS and R*. Chapman & Hall/CRC Biostatistics Series. CRC Press, 2016.
96. Prentice. RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med*. 1989;8(4):431-40

97. Papanikos T, Thompson JR, Abrams KR, Städler N, Ciani O, Taylor R, et al. Bayesian hierarchical meta-analytic methods for modeling surrogate relationships that vary across treatment classes using aggregate data. 2020;1103–24.
98. Welton, N.J., Phillippo, D.M., Owen, R., Jones, H.J., Dias, S., Bujkiewicz, S., Ades, A.E., Abrams, K.R. DSU Report. CHTE2020 Sources and Synthesis of Evidence; Update to Evidence Synthesis Methods. March 2020
99. Spiegelhalter. DJ. Bayesian graphical modelling: a case-study in monitoring health outcomes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 1998;47(1):115.
100. Lloyd S, Bujkiewicz S, Wailoo AJ, Sutton AJ, Scott D. The effectiveness of anti-TNF- α therapies when used sequentially in rheumatoid arthritis patients: A systematic review and meta-analysis. *Rheumatology*. 2010;49(12):2313–21.
101. O'Brien WA, M PH, Martin D, Esinhart J, Hill A, Benoit S, et al. Changes in plasma HIV-1 RNA and CD4+ lymphocyte counts and the risk of progression to AIDS. *N Engl J Med*. 1996;(January 1987):426–31.
102. Mellors JW, Munoz A, Giorgi J V, Margolick JB, Tassoni CJ, Gupta P, et al. Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. *Ann Intern Med*. 1997;126(12):946.
103. Sormani M, Li D, Bruzzi P, Stubinski B, Cornelisse P, Rocak S, et al. Combined mri lesions and relapses as a surrogate for disability in multiple sclerosis. *Neurology*. 2011;77(18):1684.
104. Pozzi L, Schmidli H, Ohlssen DI, Wiley J. A Bayesian hierarchical surrogate outcome model for multiple sclerosis. *Pharm Stat*. 2016;15:341–8.
105. GJ Melendez-Torres, Peter A, Xavier A, et al. Clinical effectiveness and cost-effectiveness of beta-interferon and glatiramer acetate for treating multiple sclerosis: systematic review and economic evaluation. *NIHR Journals Libr Heal Technol Assess*. 2017;21(52).
106. Dias S, Sutton AJ, Welton NJ, Ades AE. NICE DSU Technical Support Document 6: Embedding evidence synthesis in probabilistic cost-effectiveness analysis: software choices. 2012, available from <http://www.nicedsu.org.uk>.
107. Hong H, Chu H, Zhang J, Carlin. BP. A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons. *Res Synth Methods*. 2016;7(1):6–22.
108. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004;23(20):3105–24.
109. Jackson D, Bujkiewicz S, Law M, Riley RD, White IR. A matrix-based method of moments for fitting multivariate network meta-analysis models with multiple outcomes and random inconsistency effects. 2018;74(2):548–56.
110. Jackson D, White IR. When should meta-analysis avoid making hidden normality assumptions ? *Biometrical J*. 2018;60(6):1040–58.
111. Lee KJ, Thompson SG. Flexible parametric models for random-effects distributions. *Stat Med*. 2008;27(3):418–34.

112. Marshall EC, Spiegelhalter DJ. Comparing Institutional Performance using Markov chain Monte Carlo Methods. *Statistical analysis of medical data: new developments*, pages 229–249, 1998.
113. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. *Stat Med*. 1995;14(24):2685-99.
114. Roeder K, Wasserman L, Carroll RJ. Flexible Parametric Measurement Error Models. *Biometrics*. 1999;55(1):44-54.
115. Chen Y, Hong C, Ning Y, Su X. Meta-analysis of studies with bivariate binary outcomes: A marginal beta-binomial model approach. *Stat Med*. 2016;35(1):21–40.
116. Trikalinos TA, Hoaglin DC, Schmid CH. An empirical comparison of univariate and multivariate meta-analyses for categorical outcomes. *Stat Med*. 2014;33(9):1441–59.
117. White IR. Multivariate random-effects meta-analysis. *The Stata Journal*. 2009;9(1):40–56.
118. Gasparrini. A. Package 'mvmeta', <https://cran.r-project.org/web/packages/mvmeta/mvmeta.pdf>. 2018.
119. Polanin JR, Hennessy EA, Tanner-Smith EE. A Review of Meta-Analysis Packages in R. *Journal of Educational and Behavioral Statistics*, 42(2):206–242, 2017.
120. Lu G, Kounali D, Ades AE. Simultaneous multioutcome synthesis and mapping of treatment effects to a common scale. *Value Heal*. 2014;17(2):280–7.
121. Nicky J, Welton Jeroen P, Jansen Alexander J, Sutton So_a Dias AEA. *Network meta-analysis for decision making* . John Wiley & Sons; 2018.
122. Elston J, Taylor RS. Use of surrogate outcomes in cost-effectiveness models: a review of United Kingdom health technology assessment reports. *Int J Technol Assess Health Care*. 2009 Jan ;25(1):6–13. Available from: <https://pubmed.ncbi.nlm.nih.gov/19126246/>
123. Grigore B, Ciani O, Dams F, Federici C, de Groot S, Möllenkamp M, et al. Surrogate Endpoints in Health Technology Assessment: An International Review of Methodological Guidelines. *Pharmacoeconomics*. 2020;38(10):1055–70. Available from: <https://doi.org/10.1007/s40273-020-00935-1>

付録

TSD 20 の付録プログラム B.1 では、再発寛解型の多発性硬化症のデータに対して、代替エンドポイントを評価する Daniels and Hughes モデル（4.2 節）を WinBUGS コードで実装しています。ここでは SAS および Stan で実装した場合のコード例を示します。TSD 20 では WinBUGS を用いた解析結果も示されていますので、詳細は TSD 20 の 3.7 節および付録をご参照ください。なお、プログラム B.1 では、Daniels and Hughes モデルにおける試験内相関を全て共通としたモデルを用いていますので、本実装例においても同様に行っています。また、TSD 20 に示されている結果と同様な数値が得られることは確認していますが、コードの妥当性を保証しているわけではありませんのでご注意ください。

SAS コード例：

*解析データセット（25 試験）の用意；

data inds;

*代替エンドポイント y1: $\log(\text{年率換算の再発率の比})$, se1: y1 の SE;

*真のエンドポイント y2: $\log(\text{障害進行率の比})$, se2: y2 の SE;

input y1 se1 y2 se2;

index=_n_;

cards;

```
-0.0833816 0.0813719 0 0.2036533
-0.4155154 0.0893046 -0.3364722 0.2302073
-0.210721 0.3563859 0.1278334 0.7592028
-0.3424903 0.1067276 -0.1300531 0.2210844
-0.3856625 0.1266053 -0.461035 0.258417
-0.8915981 0.1351165 -0.3629056 0.3393006
-1.07881 0.2388631 -1.665008 0.7296625
-0.9942523 0.2406259 -0.2000212 0.74131
-0.3424903 0.0706827 -0.2097205 0.1464617
-0.3856625 0.0723215 -0.315081 0.1542706
0.0487901 0.0458876 0 0.0920902
-0.3424903 0.1348356 -0.8362481 0.3022435
-0.1625189 0.1005056 0.074108 0.1935841
-1.139434 0.0751103 -0.5340825 0.124761
-0.7985078 0.0603509 -0.2318016 0.0995418
-1.171183 0.1881434 -1.060872 0.3856541
```

```

-1.514128 0.2149895 -0.9555115 0.3515206
0.0295588 0.139198 0.2962659 0.2174605
-0.8675006 0.1082371 -0.3650315 0.1504737
-0.7985078 0.1038128 -0.3105963 0.145336
-0.1392621 0.1319462 0.2087549 0.3854737
-0.356675 0.1371439 0.040822 0.3966984
-0.9942523 0.2183502 -0.4462871 0.3555565
0.0582689 0.0695049 0.0487901 0.1145532
-0.0304592 0.0704836 0.0953102 0.1134564
;
run;

```

```

proc mcmc data=inds dic diag=mcse outpost=outpost stats(percentage=(2.5 50
97.5))=(sum) nbi=50000 nmc=2000000 thin=40 seed=777
monitor=(lambda20 lambda21 psi2_sq );
  array y[2] ;
  array ar_delta[2] ;
  array sigma[2,2];
  array delta1_[25];
  parms lambda20 0 lambda21 0 psi2 0.25 rho_w 0.25;
  *lambda20:  $\lambda_0$  0 に近いほど代替性がよい;
  *lambda21:  $\lambda_1$  0 だと予測に寄与していない;
  *psi2_sq:  $\psi_2^2$  試験間変動;
  *rho_w: 試験内相関 (共通) ;

  prior lambda20~normal(0,var=1000) ;
  prior lambda21~normal(0,var=1000) ;
  prior psi2~uniform(0,2);
  psi2_sq = psi2**2;
  prior rho_w ~uniform(0,0.999); * positive within-study correlation;

  sigma[1,1] = se1**2;
  sigma[2,2] = se2**2;
  sigma[1,2] = se1*se2*rho_w;
  sigma[2,1] = sigma[1,2];

```



```

    *Daniels and Hughes formulation for the between-studies model
prior ステートメントでは推定が上手くいかなかったため、random ステートメントと配列によ
って表現しています;
  parms delta1_: -0.5;
  prior delta1_ ~ normal(0, var=1000) ;
  random ar_delta2 ~ normal(lambda20+lambda21*delta1_[index], var=psi2_sq)
SUBJECT= _OBS_ INITIAL=-0.5;
  ar_delta1=delta1_[index];
  model y ~ mvn(ar_delta, sigma);*within-study model;
run ;

```

Stan コード例：

```
data {
  int num; //n of study
  //代替エンドポイント y1: Log(年率換算の再発率の比), se1: y1 のSE
  //真のエンドポイント y2: Log(障害進行率の比) , se2: y2 のSE
  vector[2] Y[num];
  vector[num] se1;
  vector[num] se2;
}

parameters {
  real<lower=0, upper=0.999> rho_w; //positive within-study correlation
  real lambda20;
  real lambda21;
  real<lower=0, upper=2> psi2;
  matrix[2,num] delta;
}

transformed parameters {
  matrix[2, 2] sigma[num];
  //covariance matrix for the i-th study
  for (i in 1:num) {
    sigma[i, 1, 1] = pow(se1[i], 2);
    sigma[i, 2, 2] = pow(se2[i], 2);
    sigma[i, 1, 2] = se1[i] * se2[i] * rho_w;
    sigma[i, 2, 1] = se1[i] * se2[i] * rho_w;
  }
}

// Bivariate model for surrogacy:
model {
  for (i in 1:num) {
    Y[i] ~ multi_normal(delta[,i], sigma[i]); //within-study model
  }
  delta[1, ] ~ normal(0, sqrt(1000));
  delta[2, ] ~ normal(lambda20 + lambda21 * delta[1, ], psi2);
}
```

```

lambda20 ~ normal(0, sqrt(1000)); //0 に近いほど代替性がよい
lambda21 ~ normal(0, sqrt(1000)); //0 だと予測に寄与していない
psi2 ~ uniform(0,2); //試験間変動
rho_w ~ uniform(0,0.999); //試験内相関 (共通)
}

generated quantities{
  real psi2_sq;
  psi2_sq = pow(psi2,2);
}

```

RStan を用いた R コード例:

Stan 用の R パッケージ RStan を用います。以下では解析データセット(1 行目を y1, y2, se1, se2 としたテキストファイル; SAS コード例の inds データセットを参照)を R にインポートし, stan への読み込み用に加工後, 実行しています。結果表示やモデル診断のコードは割愛しています。

```

dat <- read.table(<データセットのファイルを指定>, header = TRUE)
dat2 <- list(Y=as.matrix(dat[,c("y1", "y2")]),
            se1=dat[, "se1"], se2=dat[, "se2"], num=25)
library(rstan)
rstan_options(auto_write=TRUE)
options(mc.cores=parallel::detectCores())
fit <- stan(
  <stan ファイルを指定>, data=dat2, seed=1, thin=2, warmup = 2000, iter = 12000,
  control = list(adapt_delta = 0.999, max_treedepth=15)
)

```

2022 年度データサイエンス部会 KT7

作成担当者

渥美 淳	東レ株式会社
奥山 ことば	MSD 株式会社*
直井 一郎	住友ファーマ株式会社
野島 俊秋	興和株式会社

レビュー担当者

東 美恵	エーザイ株式会社
荒西 利彦	日本イーライリリー株式会社
河田 祐一	中外製薬株式会社
中島 章博	帝人ファーマ株式会社*
町田 光陽	塩野義製薬株式会社
吉田 瑞樹	ファイザーR&D 合同会社*

*TF リーダー

監事

酒井 弘憲	エーザイ株式会社
-------	----------