

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

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 2019 年度継続タスクフォース 5 「医療経済評価に関連する統計的手法の整理と紹介」

2020年12月

本報告書の中で Technical Support Document 1-7 (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、また各 TSD のすべ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、タスクフォースでの追加の解釈を加えている。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。

はじ	めに		6
1	意思	央定のためのエビデンス統合(TSD1)	9
1.1	1	費用効果分析におけるエビデンス統合の解析アプローチの概要	9
	1.1.1	ベースライン群の自然病歴(Baseline natural history)モデルと相対治療効果モデル	9
	1.1.2	エビデンス統合における意思決定と方法論	9
1.2	2	相対治療効果のエビデンス統合データセット	10
	1.2.1	比較対照薬の組(Comparator Sets)	10
	1.2.2	試験の選択・除外基準	11
	1.2.3	エビデンス統合の比較対照薬の組の拡張	12
1.5	3	エビデンスとエビデンス統合結果の表示方法	12
	1.3.1	エビデンスの提示	12
	1.3.2	エビデンスの統合方法の説明	13
	1.3.3	エビデンスの統合結果の提示	13
	1.3.4	モデル診断と選択	14
1.4	1	各 TSD で記載されている内容の要約	15
2	メタフ	アナリシス及びネットワークメタアナリシスに対する一般化線型モデルのフレームワーク	
(TSD2	2)		17
2.1	1 :	2 群比較のメタアナリシス及びネットワークメタアナリシスの導入	17
2.2	2	2 値データに対するコアモデルの展開	18
	2.2.1	2値データのメタアナリシスのロジットモデル	18
	2.2.2	モデルの規定	19
2.3	3	モデルの適合性とモデル比較	20
2.4	1	一般化線形モデル	22
	2.4.1	率データ:ポアソン尤度及び対数リンク	23
	2.4.2	率データ: 二項尤度及び cloglog リンク	24
	2.4.3	競合リスク:多項尤度及び対数リンク	25
2.5	5	間接比較への拡張とネットワークメタアナリシス	27
2.6	3	多群試験の組み込み	28

	2.7	正規尤度及び恒等式のリンク関数を伴うモデルの連続量データへの適用	29
	2.8	ベイズ流 MCMC 法における技術的な課題	33
3	異質	性:サブグループ,メタ回帰,バイアス及びその補正(TSD3)	34
	3.1	概要	35
	3.2	メタ回帰の概要	36
	3.2.1	試験内及び試験間の共変量	36
	3.2.2	生態学的誤謬(ごびゅう)(Ecologic Fallacy)	38
	3.2.3	連続値の共変量を伴う個別患者データ(IPD)を用いた群間比較でのより大きな検出力	39
	3.2.4	折りたたまれたカテゴリデータの利用(Use of collapsed category data)	39
	3.2.5	要約によるバイアス(Aggregation bias)	39
	3.3	バイアス(偏り)の調整方法の概要	40
	3.4	メタ回帰の形式としてのネットワークメタアナリシス	40
	3.5	異質性の測定	40
	3.5.1	意思決定での異質性の関わり合い	41
	3.6	外れ値の検出	42
	3.6.1	2 群比較のメタアナリシスでのクロスバリデーションによる予測の妥当性評価	42
	3.6.2	間接比較及び NMA による予測の妥当性評価のためのクロスバリデーション	48
	3.7	サブグループ,メタ回帰,ベースラインリスクへの調整	54
	3.7.1	メタ回帰の導入:サブグループ効果	54
	3.7.2	NMA における交互作用モデルの範囲とその解釈	61
	3.7.3	連続値の共変量でのメタ回帰	64
	3.7.4	ベースラインリスクによるメタ回帰	75
	3.7.5	IPD を利用したメタ回帰	80
	3.8	バイアス(偏り)とその調整	82
	3.8.1	メタ回帰による共変量の調整	83
	3.8.2	メタ疫学研究データに基づくバイアスの補正	84
	3.8.3	試験ネットワークでのバイアスの推定と調整	85
	3.8.4	専門家又はデータに基づいたバイアス分布の導出	86
4	無作	為化比較試験に基づくエビデンスのネットワークにおける不一致性(TSD4)	92

	4.1	エビデンスのループ	93
	4.2	多群試験の影響	94
	4.3	不一致性の評価方法	94
	4.3.1	Bucher 法: ひとつのループがあるネットワークへの適用	94
	4.3.2	Bucher 法の拡張: 複数のループがあるネットワークへの適用	95
	4.4	複雑なネットワークにおける不一致性の評価方法	95
	4.4.1	一致性を仮定しないモデル(Inconsistency model: Unrelated mean effects model)	96
	4.4.2	不一致性を評価するその他の方法	97
	4.4.3	複雑なネットワークにおける不一致性評価の事例	98
	4.5	不一致性への対応,異質性との関係	102
5	ベー	スラインモデルにおけるエビデンスの統合(TSD5)	105
	5.1	はじめに	105
	5.1.1	導入	106
	5.2	臨床試験アウトカムに対するベースラインモデル	107
	5.2.1	ベースラインアウトカムに関するエビデンスの情報源	107
	5.2.2	ベースライン反応に関する要約データの統合	108
	5.2.3	共変量を伴うベースラインモデル	116
	5.3	ベースラインモデルの他の部分における統合の問題	116
	5.3.1	ベースラインモデルのパラメータ及び相対的な治療効果に対する情報源	117
	5.3.2	ベースラインモデルを構成するための複数アウトカムの同時統合	118
	5.4	複数パラメータ統合によるモデルのバリデーションとキャリブレーション	119
6	確率	的な費用効果分析へのエビデンス統合の組み込み:ソフトウェアの選択(TSD6)	123
	6.1	確率論的な費用効果分析へのエビデンス統合の組み込み	123
	6.1.1	ベイズ流のシミュレーション:1段階アプローチ	124
	6.1.2	ベイズ流のシミュレーション:2段階アプローチ	125
	6.1.3	モンテカルロサンプリングによる頻度論の推定	125
	6.1.4	ブートストラップによる頻度論の推定	125
	6.2	複数のソフトウェアプラットフォームの使用	125
	6.2.1	ソウトウェアの連携	126

	6.2.2	統合プラットフォームの使用	126
7	意思	決定における治療効果のエビデンス統合:レビューアー向けチェックリスト(TSI	7) 128
	7.1	チェックリストの想定利用者	128
	7.2	想定するチェックリストの利用場面	128
	7.3	チェックリストに含まれるもの	128
	7.4	NMA における重要な仮定	129
	7.5	対比較のメタアナリシスから NMA への拡張で追加となる課題	129
	7.6	他のチェックリストとの関係	129
	7.7	チェックリストを提供する目的	129
	7.8	チェックリスト	130
8	NIC	E DSU TSD の WINBUGS プログラムについて	134
	8.1	プログラムの紹介	134
	8.2	WINBUGS の基本的な使い方の補足	135
	8.2.1	使用するデータセットの設定について	135
	8.2.2	初期値の影響があるイタレーション(反復)の破棄がある場合について	137
9	NMA	A を実施するソフトウェアの紹介	138
	9.1	NMA を実施するソフトウェアの分類	138
	9.2	MCMC 専用ソフトウェア	138
	9.3	NMA 専用のソフト	139
	9.4	汎用の統計解析ソフトウェア	140
ぉ	わりに		142

はじめに

医療経済評価を保険償還制度や価格調整制度に活用している国はヨーロッパを中心に数多くある.特にイギリスでは、早期に医療技術の費用対効果評価を導入し、医療技術評価(Health Technology Assessment; HTA)機関である NICE(National Institute for Health and Care Excellence)が標準的治療や処方を提言するガイダンスを発行している[1]. 本邦においても 2019 年 4 月、薬価の価格調整に費用対効果評価が本格導入された. HTA で主たる役割を担う費用対効果評価とは、対象となる新医療技術とそれが最も置き換わり得る医療技術(比較対照技術)とを「費用」及び「効果」の両面において比較・評価するもので、その方法論は統計学の理論に立脚しているものが多い。その中でも、薬剤効果の対照薬剤との比較において、システマティックレビューに基づくメタアナリシスの重要性が増している。また費用対効果評価に限らず、近年エビデンスに基づく医療(Evidence Based Medicine)の提供が求められ、その根拠となる学術論文のシステマティックレビュー、及びメタアナリシスの重要性が高まってきている.

メタアナリシス報告の質を向上させるために、国際研究グループにより1996年に「QUOROM(メタ アナリシス報告の質)声明」という方針が作成され、2009年にその改訂版である「PRISMA(システマ ティックレビュー及びメタアナリシスのための優先的報告事項)声明」が作成された[2].システマ ティックレビューは1つのテーマに関して、体系的かつ明示的な方法を用いて、関連研究の特定、選択 及び批判的吟味を行い、レビューに採用された研究からデータを収集・解析する研究手法である. メタ アナリシスは、システマティックレビューの結果を統合するために非常に有用な統計手法とされている. QUOROM から PRISMA への発展におけるシステマティックレビューの概念的変化として,次の4点が 挙げられている。それらは、①システマティックレビューの過程において、そのプロトコル(研究設計 書)を必要に応じて修正する必要があること、②システマティックレビューを行う過程で、採用した研 究の偏り(バイアス)のリスクに関する評価が報告されていない場合、システマティックレビューの実 施が適切に行われたとは見なされないこと、③研究レベルにおける評価(例えば、十分な症例数に基づ くものか),及び研究の特徴に由来する結果レベルにおける評価(例えば,データのバラツキ),といっ たバイアスのリスクの評価が必要であること、④報告バイアス(例えば、ポジティブな結果が発表され やすいことによるバイアス)の重大性を十分考慮すべきとされていること、である. PRISMA 声明は27 項目のチェックリスト及び 4 段階(特定,選抜,適切性,採用)のフローチャートで構成されており, システマティックレビュー及びメタアナリシスを報告する際に考慮すべき基本的事項が含まれている. さらに 2015 年にはネットワークメタアナリシス (Network Meta-Analysis; NMA)に対する PRISMA (PRISMA-NMA) [4]及びチェックリストが追加されており、本書では割愛するが NMA を実施する際に は併せて参照して頂きたい. 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[3]に おいても、PRISMA 声明の原則に従いシステマティックレビューを正しくに実施することが言及されて いる. 本邦における薬価制度への本格導入にて求められる費用対効果評価においても、システマティッ クレビューでの十分な検討が必要となることは強調されるべきであろう.

薬剤効果の比較において、適切な比較対照の薬剤と直接比較している試験が存在しない場合でも、間接比較により薬効の差を推測できる場合がある。例えば、臨床試験から治療 Avs 治療 B及び治療 Avs

治療 C の結果が得られているとき、これらの結果から直接比較していない治療 B vs 治療 C の結果を間接的に推測することである。さらに、NMA により直接比較と間接比較を含んだより包括的な薬剤効果の比較を行うことができる。NMA は、複数の治療法を比較できるという非常に魅力的な特徴をもつ反面、その結果の妥当性を担保するために満たすべき仮定が存在するため、NMA を活用する際には留意すべき事項が追加される。先に記載したシステマティックレビューの過程においても、試験間の異質性又はネットワークでの不一致性を説明できる因子が示唆されたならば、その因子をモデルに反映することも可能であり、また結果の解釈でも重要な示唆を与えると考えられる。そのため、統計解析担当者にとっては、これらの方法論の正確な理解、及び結果の慎重な解釈が必要となる。

NICE では、2011 年から 2012 年に NMA に関連した 7 つの Technical Support Documents (TSD) [TSD1~TSD7] を発行している. これらの TSD は、費用対効果評価に関わる意思決定のためのエビデンス統合の手法における技術サポートの基礎的な内容を含んだシリーズであり、費用対効果評価に用いる相対的な治療効果(参照治療に対する治療効果)は、無作為化比較試験(Randomized Controlled Trial; RCT)の統合に焦点があてられている。様々な評価指標に対応できる一般化線形モデルの枠組みにより NMA の手法が紹介されており、NMA にて共変量などを考慮するメタ回帰モデル、それらの頑健(ロバスト)性を脅かす異質性又は不一致性の検討及び回避方法について、WinBUGS による解析事例を含めて具体的に説明されている。さらに、WinBUGS を始めとする利用可能なソフトウェア、及び NMA による治療効果のエビデンス統合結果についてレビューアーがチェックすべき項目のリストが提示されている。費用対効果評価での費用算出においては、NDB等のリアルワールドデータが活用されているものの、効果のエビデンス統合(対照薬との相対的な効果の大きさ)に関しては、エビデンスレベルの高い RCT データによる評価が原則となっている[3]. 一方、RCT が存在しない又は RCT の結果が実際の臨床成績と大きく乖離しているような場合は、製造販売業者と国立保健医療科学院/公的分析班等との協議の上、他の RCT を用いる又は観察研究を用いるなどの方法も取り得る.

本報告書は、本邦において HTA を用いた申請を検討している実務者が、他剤を対照として自社の薬剤の効果を比較するために NMA を活用する際の基本的な知識を得ることを目的としている。統計担当者以外の実務者にも幅広く理解いただけるよう、数式については可能な限り平易に説明を行い、多くの事例を盛り込んでいる。HTA における NMA の活用を一通り理解いただけるよう、TSD1~7 のそれぞれについて、特に重要と考えられる部分を中心に解説する。なお、ベイズ統計に馴染みのない方は、必要に応じて文献 [5][6] 等を参照されるとより理解が深まると考えられる。また 1.4 節に各 TSD の概要を記載しているので、まずはそちらを参照することをお勧めする。TSD2~4 で紹介されている WinBUGS プログラムコードについては SAS プログラムに移植を行い、SAS ユーザー総会の Web site にて発表資料に加えて SAS プログラム (追加資料) のダウンロードが可能であるので、適宜ご参照頂きたい[7]。

参考文献:

- [1] National Institute for Health and Care Excellence. Guidance List. URL: https://www.nice.org.uk/guidance/published
- [2] 卓興鋼, 吉田佳督, 大森豊緑. エビデンスに基づく医療 (EBM) の実践ガイドライン, システマティックレビューおよびメタアナリシスのための優先的報告項目 (PRISMA 声明). 情報管理. 2011; 54: 254-266.

- [3] 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2版. URL: https://c2h.niph.go.jp/tools/guideline/guideline_ja.pdf
- [4] PRISMA for Network Meta-Analyses (PRISMA-NMA) http://www.prisma-statement.org/Extensions/NetworkMetaAnalysis.aspx
- [5] 久保 拓弥. データ解析のための統計モデリング入門 一般化線形モデル・階層ベイズモデル・MCMC(確率と情報の科学). 岩波書店. 2012
- [6] 繁桝 算男. ベイズ統計入門. 東京大学出版会. 1985
- [7] 奥山ことば,町田 光陽, 石渡 量太, 渥美 淳, 奥村 広之. 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. SAS ユーザー総会 2019

URL: https://www.sas.com/ja jp/events/19/users-group-2019/sessions.html#m=15

1 意思決定のためのエビデンス統合(TSD1)

Technical Support Documents(TSD)は、費用対効果評価における意思決定のためのエビデンスを統合する方法として、モデルを用いた解析手法について紹介している. TSD は英国の医療技術評価機関 NICE (National Institute for Health and Care Excellence; NICE)[1]への申請資料をサポートすることを目的としており、どの分析モデルを用いるか等を規定するのではなく、必要な特性を満たす分析モデルの要件やその選択方法について紹介している. TSD1 では評価の対象とする試験は介入試験であり、主に対象とする試験の相対的および絶対的な有効性の結論について取扱い、エビデンス統合のプロセスについての一般的なガイダンスを提供している.

1.1 費用効果分析におけるエビデンス統合の解析アプローチの概要

1.1.1 ベースライン群の自然病歴(Baseline natural history) モデルと相対治療効果モデル

一般的な費用対効果評価(Cost Effective Analysis; CEAs)モデルは、2つの要素から構成され、1つは標準治療(対照群)のもとでの対象となる疾患に新たに罹患する患者の自然病歴(natural history:現行治療での治療経過)に対するベースラインモデルと、もう1つは治療群での追加的に得られる相対治療効果に対するモデルである。治療群における自然病歴は、ベースラインモデルと相対治療効果モデルを適切に組み合わせることで絶対効果を得る。ベースラインモデルにおける相対効果が作用すると考えられる時点は、モデルにより異なる。例えば、関節リウマチの治療効果モデルでは、少なくとも患者が3~6ヶ月間は治療を継続する確率に基づいており[2][4]、禁煙のモデルでは治療期間(12ヶ月以内)に再び喫煙してしまう確率に基づいている[5][6]。このような自然病歴モデルの(不変的な特徴ではないが)共通した特徴は、初期の短期的効果を条件付けたもとで、中長期的なアウトカムが治療に依存しない仮定をおくことである。マルコフモデルを用いる場合、治療効果(相対及び絶対効果を含む)は1つまたは複数の推移確率に従って推移すると仮定する。ベースラインと相対治療効果を推定するためのデータソースを特定するために、システマティックレビューを用いるべきとされている[1]。TSD5[7](本書第5章)ではベースラインモデル構築方法、TSD2[10](本書第2章)では相対効果の統合方法について紹介されている。

1.1.2 エビデンス統合における意思決定と方法論

TSD1~TSD7 で提案されている方法は、確率論的モデリング(例えば、マルコフモデル)に反映するために設計されている。どのような分析モデルや解析ソフトウェアを選択するということではなく、費用対効果評価の意思決定の問題を明確に定義し、エビデンスを統合するあらすじを組み立てることが重要になる。分析対象となる集団は、ある時点の特定の状態だけを考慮するだけでなく、経時的な疾患の進展過程なども考慮して構成する。また候補となる介入試験の選択には、用法用量・投与経路、併用療法などを詳細に取り決める必要がある。費用効果分析モデルにより、各治療群で1つの効果と、1つの費用を計算する必要があり、費用は疾患の進展過程に依存し経時的に変動する可能性があるため、適切に平均化される必要があり、効果も同様に変化する可能性があるため、モデル化が必要となる。

費用対効果評価、又は有効性に関する解析のみであったとしても、エビデンス統合に含める試験に対する選択基準は、文献要約のためのシステマティックレビューでの選択基準より狭くなりやすいことが知られている[8][9]. そのため、異質性 (heterogeneity) を平均化する変量効果モデル (random effect model) を用いる場合は十分な考察を必要とする. 同様の観点からコクランハンドブック[11]では、「メタアナリシスは、意味のある要約を提供するためには、試験の対象患者、介入およびアウトカム (評価指標) に関して、試験間で十分に均質な場合にのみ考慮されるべきである.」としている. 例えば、慢性関節リウマチ治療の場合、標準療法が効かない患者を対象とした生物製剤の試験と、生物製剤試験で効果が得られなかった患者の試験を組み合わせでは、生物製剤の相対的な有効性は評価できない[12]等である. 臨床的に異質な患者集団に対しエビデンスを統合する場合、統計的異質性 (heterogeneity) と不一致性 (inconsistency) のリスクを増大するだけでなく、介入に対しナイーブな集団とナイーブでない集団で同様に効果的であるといった、臨床現場では受け入れられない誤った仮定を必要とする可能性があるため、十分に留意する必要がある.

1.2 相対治療効果のエビデンス統合データセット

分析対象集団は、疾患の進行ステージ・重症度、脱落理由、前治療もしくはこれらの組み合わせ等を 参照して定義する. 対象集団が広くなると、臨床的および統計的異質性が高まり、別グループに属する 対象集団に対し、誤った意思決定することになる.

1.2.1 比較対照薬の組(Comparator Sets)

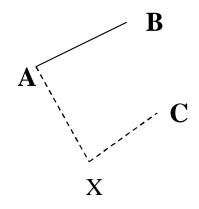


図 1-1 意思決定のための比較対照薬の組(A, B, C)とエビデンス統合のための比較対照薬の組(A, B, C, X)が異なるネットワーク。実線又は破線は、少なくとも 1 つの試験において、ネットワークを構成する治療の比較があったことを示している。意思決定に関連する治療は太字で示す。治療 X は治療 C をネットワークの残りの部分(破線)にリンクさせることができるため、エビデンス統合のための比較対照薬の組に追加する。

[NICE DSU TSD1 の Figure 1 を改変して掲載]

1.2.2 試験の選択・除外基準

分析対象集団とエビデンス統合のための比較対照薬の組をそれぞれ定義してしまえば、試験を選択する基準を適用するのは比較的簡単である。また、問題としている患者集団において、エビデンス統合の比較対照薬の組に属する 2 つ以上の治療薬を比較している全ての試験を組み入れるべきである。これは、特定の結果を得るために意図して試験を選択、または除外するリスクを減らすためである。例えば、3 つの治療(A-B-Z)を比較する 1 つの試験があり、治療 Z はエビデンス統合のための比較対照薬の候補ではない場合、それ以上の情報を追加しないので除外されるかもしれないが、同じ患者対象集団である限り、治療 A と治療 B の比較の情報は用いるべきである。エビデンス統合のための比較対照薬の組に含まれる治療法が比較されている全ての試験を組み入れることが重要である。その理由として、意思決定は増分費用効果分析(Incremental CEA)に基づいており、治療(A, B, C)に関して、試験(A-B)、(A-C)、(B-C) はすべて同程度に有益な試験であるためである。実際には、「直接的」および「間接的」エビデンスを区別すること(すなわち 1 つのモデルパラメータの直接的エビデンスが別のモデルパラメータの間接的エビデンスであるかを区別すること)は、技術的に難しいことがある。

ランダム化は行っているが二重盲検(double-blind)ではない試験,あるいは症例数の少ない試験を除くべきとの議論があるが,一般的に上記の規則に該当する場合,どの試験も明確な理由がない限り除外すべきでない.例えば,症例数の少ない試験を除外する理由として考えられるのは,バイアスが生じている可能性が高い場合である.このバイアスは小規模試験バイアス(small study bias)と呼ばれる.これは「試験規模」が,試験の質,あるいは効果が(偶発的に)大きくなる要因の「代替的な指標」として見なされることがあるためである.この場合,CEA は規模の小さい試験の選択・除外の基準に対してどのくらい頑健性(感度)があるか,バイアスを調整する方法がどのような影響を及ぼすかを検討すべきである(TSD3[14]).また,非劣性試験というだけで除外することは妥当ではない.非劣性試験での症例数(同じ実薬対照との優越性試験より症例数が少ない場合がある)が,実際に得られる治療効果の大き

さ,又はその方向性との間には関連性がないためである(つまり,設定した症例数が,治療効果に影響を及ぼすことはない).その他に試験を選択しない理由として,アウトカムが他の試験とは異なる形式で報告されているような場合である.常に可能とは限らないが,場合によっては変換可能な場合もあり,TSD2[10]で幾つかの事例を取扱っている.アウトカムが統合可能そうであったとしても,試験の選択・除外の基準について,試験の対象集団が意思決定の対象集団と一致するか,エビデンス統合のための比較対照薬で定義された治療法と,当該試験で検討した治療法が十分に類似しているのか等の主観的な判断は必要となる.

1.2.3 エビデンス統合の比較対照薬の組の拡張

ネットワークを拡張できる範囲は、ネットワークを生成するための拡張が必要でなければ、エビデン ス統合の比較対照薬の組の範囲を超えてまで拡張はする必要はない(例えば、ある疾患領域において、 幾つもの薬剤が利用可能であったとしても、当該の意思決定に含まれない薬剤は含める必要がない). ネットワークを拡張する利点として、推論 (inference) の強化[15]、特定のデータに依存しない頑健 (ロ バスト)な解析結果が得られること、評価対象のエビデンスが増えることでより一致性(consistency) の評価が安定することを確認できる(TSD4[16]),効果修飾因子(effect modifiers)を同定しバイアスを 調整できる(TSD3[14]参照)等がある. ただし, 潜在的な欠点は, かけ離れている治療法をつなげるこ とで、多少異なる患者集団までも含めることになり、新たな効果修飾因子を導入してしまう可能性を高 めることである. 最後に、エビデンス統合の比較対照薬の組を拡張する別の理由は、エビデンスの欠如 を埋め合わせるために、アウトカム間の関係性を推定できるということが挙げられる。事例として、癌 の費用効果分析モデルには奏功割合、増悪期間、および生存期間に関する情報を必要とするが、これら の一部の情報はエビデンス統合の比較対照薬の組からは得られないことがある. そのような場合, より 広い範囲の試験を組み合わせることで、その関係性を推定でき、また欠如している治療効果に対して、 推定した関係性を適用できることがある.複雑なエビデンス統合の利点としては、データ数が疎ら (sparse) であるとき、より安定した推定値を与えること、あるいはアウトカムがあまり一般に報告され ていない場合でも推定値を与えることができる等が挙げられる.ただし,さまざまな前提条件が必要と なり、特にアウトカム間での特定の関係性の不変性については、可能な限り専門家の臨床的意見と照合 する必要がある.

1.3 エビデンスとエビデンス統合結果の表示方法

1.3.1 エビデンスの提示

複数の比較対照治療を比較するネットワークダイアグラム (network diagram) を図 1-1 に示した. このような図を描くソフトウェアは R パッケージやその他プログラム等を利用でき、一般的には、比較する試験の数を実線に併記し、実線の太さで対象となる試験数を表現し、ノード (node; 節点) の大きさで当該治療法の患者数を示すことが多い.表 1-1 のように、試験ごと比較した治療群の結果を示し、試験レベルでの共変量を追加したり、個々の被験者データ (IPD) の利用可能性などを示したりするのも有益である (TSD3[14]).

表 1-1 慢性関節リウマチ治療の事例

Study Name	Placebo	CZP	Adalimumab	Infliximab	Rituximab	Mean disease duration (years)
Kim 2007	9/63		28/65			6.85
DE019	19/200		81/207			10.95
RAPID 1	5/62	146/393				6.15
START	33/363			110/360		8.1
Strand 2006	5/40				5/40	11.25

各治療群における 6ヵ月時点の ACR50 達成数/患者数 [NICE DSU TSD1 の Table 1 を改変して掲載]

1.3.2 エビデンスの統合方法の説明

解析手法やデータの透明性を高めるため、費用対効果評価のレビューアーに十分な情報(もし可能であれば、解析結果の再現に必要なデータへのアクセス、仮定した正確なモデル及びソフトウェアのソースコード/文献引用等)を提供しなければならない、機密性の制限により公開できない場合もあるが、使用されたデータと供に、エビデンス統合に用いたソフトウェアコードには注釈を付与し、レビューアー利用可能にする必要がある。モデルに組み込んだ統計的および臨床的仮定、最終的な意思決定についての考察を記載し、特定の指標(例:オッズ比、ハザード比、リスク差など)でアウトカムのモデルを選択した理由と相対的な関係から絶対的な効果への変換で必要となった仮定を明記する必要がある(TSD2[10])。

1.3.3 エビデンスの統合結果の提示

費用効果分析(CEA)の結果について、治療間の相対効果がどのように想定されたのか、絶対効果がCEAにどのように組み込まれたのかを示す必要がある。以下にパーキンソン病の補助療法の事例を示す。4種類の治療法(ドーパミンアゴニスト)とプラセボを投与された患者集団での相対的および絶対的な平均 off time(症状がコントロールできていない期間)の短縮期間を表 1-2 に、治療ネットワークを図 1-2 に示す。

表 1-2 パーキンソン病の補助療法

		mean	sd	95% CrI(信用区間)			
X	Y	Relative e	Relative effects of treatment Y compared to X				
Placebo	Treatment 2	-1.81	0.33	(-2.46,-1.16)			
Placebo	Treatment 3	-0.47	0.49	(-1.43,0.49)			
Placebo	Treatment 4	-0.52	0.48	(-1.46,0.43)			
Placebo	Treatment 5	-0.82	0.52	(-1.84,0.22)			
		Absolute treatment effects					
Plac	cebo	-0.73	0.22	(-1.16,-0.30)			
Treati	ment 2	-2.54	0.40	(-3.32,-1.76)			
Treati	ment 3	-1.21	0.53	(-2.25,-0.15)			
Treati	ment 4	-1.25	0.53	(-2.28,-0.21)			
Treati	ment 5	-1.55	0.57	(-2.66,-0.43)			

[NICE DSU TSD1 の Table 2 を改変して掲載]

1.3.4 モデル診断と選択

意思決定においては、得られたモデルが内的に一貫しており、かつ利用できるすべてのエビデンスとの整合性が取れていることが重要になる。その評価には、モデルの適合度の評価、異質性および外れ値の検出、治療効果の統合に用いた固定効果(Fixed effect)又は変量効果(Random effect)モデルの選択の説明、複数のモデルが候補となった場合にモデル選択が費用効果分析へ影響するか検討するための感度分析などが必要となる。

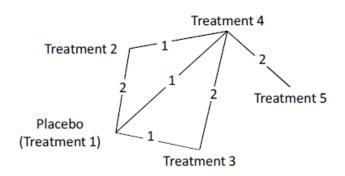


図 1-2 パーキンソン病の補助療法ネットワーク

各ノードは治療を表し、ノードとノードを接続する実線は無作為化試験で直接比較された治療法の対を示す. 実線上の数字はその比較が行われた試験数を示す.

[出典: NICE DSU TSD1 Figure 3]

1.4 各 TSD で記載されている内容の要約

TSD2[10]では通常の対比較におけるメタアナリシスとネットワークメタアナリシス(Network Meta-Analysis; NMA)の一般的な概要を示し、データタイプ、アウトカムの種類、前提条件、および分析に用いる固定効果と変量効果モデルの詳細を示す。データタイプに適切なモデルを用いる重要性、異なるフォーマットで得られたデータをどのように組み合わせるかを示す。TSD3[14]では外れ値の検出、共変量又はバイアスの調整について検討する。対比較と NMA における回帰モデルは、メタ回帰の前提条件とその危険性、ならびに意思決定での適用方法を検討する。TSD4[16]では、NMA における不一致性(inconsistency)を定義、その検出方法、回避すべき原因とその方法を示す。TSD5[7]では、比較対照群におけるベースラインモデル(baseline model)を構築するガイダンス、そのためのエビデンスの検討方法、ベースラインモデルの事例を示す。TSD2、3、4、5[10][14][16][7]に WinBUGS の Bayesian マルコフ連鎖モンテカルロ(MCMC)法[17]を用いた解析の事例と使用したプログラムコードが提供されている。TSD6[13]では、TSD2~TSD5で用いたモデルを実装できるソフトウェア、エビデンスの統合結果を意思決定モデル(補足:状態推移モデルなどの経済モデル)に組み込む幾つかのソフトウェアを紹介する。最後に、TSD7[18]ではエビデンス統合に関するレビューアーのチェックリストを提示し、統合で必要となる仮定、仮定を裏付ける根拠の妥当性を明らかにする等、様々な項目に焦点を当て、さらなる解析の必要性を決定するのに用いることができる。

本報告書で紹介した TSD1~TSD7 の著者らは、TSD の内容に加えて、事象が発現するまでの時間(time to event) や複数のアウトカムの統合などの発展的な内容を追記した書籍[19]を執筆している. TSD の内容の補足説明や事例の追加さらには一部誤記訂正も行われているため、TSD を読む際には大変参考になると思われるため、併せて参照して頂きたい.

参考文献:

- [1] National Institute for health and Care Excellence. Guide to the methods of technology appraisal 2013.
- [2] Brennan, A., Bansback, N., Nixon, R., Madan, J., Harrison, M., Watson, K. et al.
- [3] Modelling the cost effectiveness of TNF-alpha antagonists in the management of rheumatoid arthritis: results from the British Society for Rheumatology Biologics.Registry. Rheumatology 2007; 46:1345-1354.
- [4] Woolacott, N.F., Bravo Vergel, Y., Hawkins, N., Kainth, A., Khadjesari, Z., Misso, K. et al. Etanercept and infliximab for the treatment of psoriatic arthritis: a systematic review and economic evaluation. Health Technology Assessment 2006; 10(31).
- [5] Woolacott, N.F., Jones, L., Ford, C.A., Mather, L.C., Sowden, A.J., Song, F.J. et al. The clinical effectiveness and cost-effectiveness of buproprion and nicotine replacement therapy for smoking cessation: a systematic review and economic evaluation. Health Technology Assessment 2002; 6(16):1-245.
- [6] Wang, D., Connock, M., Barton, P., Fry-Smith, A., Aveyard, P., Moore, D. 'Cut down to quit' with nicotine replacement therapies in smoking cessation: a systematic review of effectiveness and economic analysis. Health Technology Assessment 2008; 12(2).
- [7] 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
- [8] Eccles, M., Freemantle, N., Mason, J. Using systematic reviews in clinical guideline development. In: Egger M., Davey Smith G., Altman D.G., eds. Systematic Reviews in Health Care: meta-analysis in context. Second ed. BMJ; London: 2001; 400-409.
- [9] Rubin, D.B. A new perspective. In: Wachter K.W., Straf M.L., eds. The Future of Meta- analysis. Russell Sage Foundation; New York: 1990.
- [10] Dias, S., Welton, N.J., Sutton, A.J., Ades, A.E. NICE DSU Technical Support Document 2: A generalised linear modelling framework for pair-wise and network meta-analysis of randomised controlled trials. 2011; last updated April 2012; available from http://www.nicedsu.org.uk
- [11] Higgins, J.P.T., Green, S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.0 (updated February 2008). The Cochrane Collaboration, Wiley, Chichester; 2008.
- [12] Wells, G.A., Sultan, S.A., Chen, L., Khan, M., Coyle, D. Indirect Evidence: Indirect Treatment Comparisons in Meta-Analysis. Cochrane Database of Systematic Reviews 2009. Ottawa, Canadian Agency for Drugs and Technologies in Health.
- [13] Dias, S., Sutton, A.J., Welton, N.J., Ades, A.E. NICE DSU Technical Support Document 6: Embedding evidence synthesis in probabilistic cost-effectiveness analysis: software choices. 2011; last updated April 2012; available from http://www.nicedsu.org.uk
- [14] 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
- [15] Cooper, N.J., Peters, J., Lai, M.C.W., Juni, P., Wandel, S., Palmer, S. et al. How valuable are multiple treatment comparison methods in evidence-based health-care evaluation? Value in Health 2011; 14:371-380.
- [16] 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 2012; available from http://www.nicedsu.org.uk
- [17] 丹後俊郎, Taeko Becque. ベイジアン統計解析の実際~WinBUGS を利用して~. 朝倉書店. 2011.
- [18] Ades, A.E., Caldwell, D.M., Reken, S., Welton, N.J., Sutton, A.J., Dias, S. NICE DSU Technical Support Document 7: Evidence synthesis of treatment efficacy in decision making: a reviewer's checklist. 2012; available from http://www.nicedsu.org.uk
- [19] Dias, S., Ades A.E., Welton N.J., et al. Network Meta-Analysis for Decision-Making. John Wiley & Sons, Inc. 2018.

2 メタアナリシス及びネットワークメタアナリシスに対する一般化線型モデルのフレームワーク (TSD2)

TSD2 では、無作為化比較試験 (Randomized Controlled Trial; RCT) から得られたデータの統合における一般化線形モデルのフレームワークを提供している。また、固定効果モデルと変量効果モデルのそれぞれに対する線形回帰に基づく共通のモデルを与え、その中核となるモデル(以下、コアモデルと呼ぶ)をネットワークメタアナリシスに適用している。推定ではベイズ流アプローチを適用し、Markov Chain Monte Carlo (MCMC) 法によるシミュレーションのための WinBUGS プログラムコードも与えている。各モデルの適合度は残差逸脱度に基づく情報量規準 (Deviance Information Criteria; DIC) により比較されている。最後に、MCMC における計算上の課題への対応について提案している。

2.1 2 群比較のメタアナリシス及びネットワークメタアナリシスの導入

メタアナリシスは独立な情報源,特にRCTからのエビデンスを統合するための手法として今日,医学研究文献において一般的に行われている.この統計手法には,2×2分割表からの結果を併合する手法を含み,多くの文献で紹介されている.メタアナリシスの幅広い適用には,長年にわたる統計手法やソフトウェアの発展が寄与しており,Sutton & Higgins[1]により最近の発展における展望が与えられている.

TSD2 では、確率論的な意思決定(決定論的ではなく確率論的な情報に基づく意思決定)に活用できるように、RCT から集積されたエビデンスを統合するための統一的な解釈及び説明を与える.臨床試験から報告される様々な形式のデータを線形に変換することにより(補足:より正確には反応変数の期待値をリンク関数により変換したものに対し、線形モデルを構築する)、一般化線形モデルの枠組みを適用する.さらに、正規分布に基づく尤度(以下、正規尤度と呼ぶ)、二項分布に基づく尤度(以下、二項尤度と呼ぶ)、ポアソン分布に基づく尤度(以下、ポアソン尤度と呼ぶ)、多項分布に基づく尤度(以下、多項尤度と呼ぶ)について、恒等式(identity)、Logit 変換、Log 変換、complementary log-log 変換、probit 変換のリンク関数を用い、固定効果モデルと変量効果モデルでの線形予測による共通のコアモデルを与える.

ネットワークメタアナリシス(Network Meta-Analysis; NMA)として知られている 3 治療以上が含まれる混合比較(Mixed Treatment Comparisons; MTC)は、エビデンス統合の比較的最近の発展によるものである。NMA は、治療 A と治療 B を比較する試験の情報を併合するだけでなく、複数の無作為比較試験からの治療 A vs 治療 B、治療 A vs 治療 C、治療 A vs 治療 D、治療 B vs 治療 D などのデータを併合し、RCT でのエビデンスを考慮しつつ、ネットワークでの内部的に一致した推定値を提示する。NMA のために構築するコアモデルは、2 群のメタアナリシス、多群試験、間接比較、及び NMA に区別なく適用できる。2 群比較のメタアナリシスや間接比較は、NMA の特別なケースである。

一般的な一般化線形モデル(Generalised Linear Model; GLM)の枠組みは、頻度論(補足:ベイズ流に

対し頻度流と呼ばれたりもするが、ここでは頻度論とする)またはベイズ流のどちらの方法にも当てはめられる. しかしながら、長年の間、包括的な決定分析(comprehensive decision analysis)の主軸は MCMC を利用するベイズ流であった. それはベイズ流の事後分布に基づくシミュレーションが統計的推定を行えるだけでなく、不確実性を考慮した下での確率的な意思決定の基盤を与えるからである. 無償で利用できる WinBUGS 1.4.3 MCMC パッケージ(ただし、現在メンテナンスは中止されている) は、エビデンス統合の GLM アプローチを実装できるモジュールを活用することで、(ネットワーク)メタアナリシスの実行及びモデル吟味に関して、固定効果モデルと変量効果モデルを統一的に扱うことができる.

2.2 2値データに対するコアモデルの展開

ここでは、事前に規定された対象患者集団において治療 1 と治療 2 を比較している M 個の臨床試験に対してメタアナリシスを実施することを想定する。固定効果モデルでは、各試験iがサンプリング誤差を伴う試験間で共通のパラメータdを推定する。一方、変量効果モデルでは、各試験iでの試験特有の治療 1 に対する治療 2 の効果 $\delta_{i,12}$ について、等しくはないが交換可能(exchangable)であると仮定して推定する。この仮定の下では、全ての $\delta_{i,12}$ が類似していることが前提になるため、試験番号i 自体は意味を持たない。この交換可能性の仮定は、試験特有の治療効果が平均 d_{12} と分散 σ_{12}^2 を伴う試験間で共通な分布に基づくことに等しい。この分布には通常正規分布が想定され、

$$\delta_{i,12} \sim N(d_{12}, \sigma_{12}^2)$$

となる。固定効果モデルはこの特別な場合で、分散をゼロに設定することにより得られる。2 つの治療のみのメタアナリシスの場合は、治療比較が 1 回のみのため、d, δ , σ の添え字は不要になる。以降では、共通の分散の仮定のもとで σ の添え字は削除するが、 δ と d の添え字は維持する。これは複数治療への拡張を考慮するためである。

2.2.1 2 値データのメタアナリシスのロジットモデル

心筋梗塞後の死亡を避けるための β 遮断薬の 22 試験のメタアナリシスを検討する. 利用可能なデータは, 22 試験における,対照群(治療 1)及び β 遮断薬群(治療 2)での患者数とそのうちの死亡数(イベント数)である(表 2-1).

表 2-1 β 遮断薬の事例: 22 試験の対照群とβ遮断薬群における患者数と死亡数

->#4€	対則	照群	β遮断薬群			
試験	死亡数 (r _{i1})	患者数 (n _{i1})	死亡数 (r _{i2})	患者数 (n _{i2})		
1	3	39	3	38		
2	14	116	7	114		
3	11	93	5	69		
4	127	1520	102	1533		
5	27	365	28	355		
6	6	52	4	59		
7	152	939	98	945		
8	48	471	60	632		
9	37	282	25	278		
10	188	1921	138	1916		
11	52	583	64	873		
12	47	266	45	263		
13	16	293	9	291		
14	45	883	57	858		
15	31	147	25	154		
16	38	213	33	207		
17	12	122	28	251		
18	6	154	8	151		
19	3	134	6	174		
20	40	218	32	209		
21	43	364	27	391		
22	39	674	22	680		

2.2.2 モデルの規定

試験 i の治療群 k について、各群の患者数 n_{ik} からイベント数 r_{ik} を定義し、イベント発生過程が二項分布に従うと仮定する. すなわち、

$$r_{ik} \sim \text{Binomial}(p_{ik}, n_{ik})$$
 (1)

であり、 p_{ik} は試験 i の治療群 k におけるイベントの発現率を示す(i=1,...,22; k=1,2).興味のあるパラメータ p_{ik} は死亡率で、0 から 1 の間の値を取るため、これらの確率を±無限大の間の連続尺度に変換するためにリンク関数が利用される.二項尤度について最も一般的に利用されるリンク関数は、 \log it リンク関数(\mathbf{z} 2-3 参照)である.ここでは \log it 尺度上での死亡率 p_{ik} を次のようにモデル化する.

$$logit(p_{ik}) = \mu_i + \delta_{i,1k} I_{\{k \neq 1\}}$$
(2)

ここに,

$$I_{\{u\}} = \begin{cases} 1 & \text{if } u \text{ is true} \\ 0 & \text{otherwise} \end{cases}$$

である.この設定において, μ_i は試験特有のベースライン値で「対照」治療(すなわち,治療 1)における死亡率の対数オッズを表し, $\delta_{i,12}$ は対照群(治療 1)に対する β 遮断薬群(治療 2)における試験特有の死亡率の対数オッズ比である.ここで式(2) を

$$logit(p_{i1}) = \mu_i$$

$$logit(p_{i2}) = \mu_i + \delta_{i,12}$$

と書き換えることができる.変量効果モデルでは、試験特有の対数オッズ比は試験共通の分布 $\delta_{i,12} \sim N(d_{12}, \sigma^2)$ に従う. 固定効果モデルでは、式(2) を

$$logit(p_{ik}) = \mu_i + d_{12} \times I_{\{k \neq 1\}}$$

で置き換える。それは、基盤となる真の治療効果の同質性を仮定して、試験間の異質性 σ^2 の設定をゼロとすることに対応する(その場合、試験間のバラツキはサンプリング誤差と解釈する)。ここで提示するメタアナリシスモデルの重要な特徴は、試験特有のベースライン値 μ_i にモデルを仮定していないことである。これはモデルで推定される局外母数 (nuisance parameters) として考慮される。ベースライン(リスク)のモデル化に関しては割愛するが、TSD5 で議論されている。

2.3 モデルの適合性とモデル比較

あてはめられたモデルの適合度が十分かどうかを点検するために、適合度の尺度を検討する.残差の逸脱度を \bar{D}_{res} で示し、モデルのあてはめを改善できるか点検するため、 \bar{D}_{res} の値と独立なデータポイントの個数(例えば、 β 遮断薬の例では、22 試験 $\times 2$ 群=44 データポイント)を比較する.

二項尤度の場合、試験毎の各群が1つの独立なデータポイントとなり、MCMC 法による繰り返し毎に計算される全データポイントの残差逸脱度の合計 D_{res} は次で計算される.

$$D_{res} = \sum_{i} \sum_{k} 2 \left(r_{ik} \log \left(\frac{r_{ik}}{\hat{r}_{ik}} \right) + (n_{ik} - r_{ik}) \log \left(\frac{n_{ik} - r_{ik}}{n_{ik} - \hat{r}_{ik}} \right) \right)$$
$$= \sum_{i} \sum_{k} dev_{ik}$$
(3)

ここに、 r_{ik} 、 n_{ik} 及び p_{ik} は試験 i の群 k での観測イベント数、患者数及びイベント発現率を表し、 \hat{r}_{ik} = n_{ik} p_{ik} は現在のモデルに基づき、それぞれ繰り返しで計算される試験毎の群 k でのイベントの期待数で、 dev_{ik} は、MCMC 法による繰り返し毎に計算される各データポイントにおける残差逸脱度である。これは全データポイントの残差逸脱度の合計 D_{res} の繰り返し計算の平均、すなわち事後平均 \bar{D}_{res} により要約される。そして各データポイントの影響度を評価するために効力統計量 (Leverage statistics) が利用できる。この統計量は頻度論における回帰分析で一般的である。各データポイントの効力である $leverage_{ik}$ は、残差逸脱度の事後平均から、事後平均での逸脱度を差し引いたものとして計算される。二項尤度については、 \hat{r}_{ik} の事後平均を \hat{r}_{ik} 、 dev_{ik} の事後平均を \overline{dev}_{ik} 、その平方根を w_{ik} と表わす。 dev_{ik} は式(3)で \hat{r}_{ik} を \hat{r}_{ik} に置き換えて計算した逸脱度である。そして、

$$p_D = \sum_{i} \sum_{k} leverage_{ik} = \sum_{i} \sum_{k} \left[\overline{dev}_{ik} - \widetilde{dev}_{ik} \right]$$
 (4)

と表すと、DIC は残差逸脱度の事後平均 \bar{D}_{res} と p_D (有効なパラメータ数とも呼ばれる) の和となる.

$$DIC = \overline{D}_{res} + p_D$$

DIC は、モデルの複雑性のペナルティ (p_D) を課したモデル適合の尺度である。DIC が小さいモデルは、より倹約したモデルであることを示唆する。特に、同じ尤度及び同じデータについて様々なモデルを比較する際には便利である。例えば、固定効果モデルと変量効果モデルの比較、固定効果モデルで共変量を含む場合と含まない場合の比較(但し、後述するように DIC は変量効果モデルに共変量を含めるかどうかの規準にはならない)などであり、3 ポイント程度の減少(増加)が適合度の向上(低下)の目安となる[2][3]。

SAS または WinBUGS がデフォルトで出力する残差逸脱度には定数部分も含められているが($DIC=\bar{D}_{model}+p_D$, $\bar{D}_{model}=\bar{D}_{res}+$ 定数部分),定数部分は共通であるためモデル間の比較をする上で定数部分は本質的ではない。ただし,WinBUGS の DIC ツールにより出力される p_D は上述した \hat{r}_{ik} からではなく, d_{ik} の事後平均から算出した残差逸脱度を用いている。パラメータの不確実性が大きく非常に非線形なモデルでは \hat{r}_{ik} を用いる方がより安定する。TSD では,leverage の図示以外は WinBUGS の DIC ツールによる p_D を用いており,恐らく WinBUGS 単体では式(4) による計算はできないためであり,TSD の結果を参照する場合は注意されたい。

表 2-1 の β 遮断薬の事例に固定効果モデル及び変量効果モデルをあてはめた結果を表 2-2 に示す. 2 つのモデルの間に DIC の差はほとんどないため、シンプルな固定効果モデルが好まれるかもしれない. 一方で効力に着目すると、固定効果モデルで大きい値を示した 2 つのポイントは試験番号 14 の 2 つの群であり、変量効果モデルではこの 2 つのポイントの効力は小さくなっている(図 2-1 に黒丸で囲われているポイント). そのため、この試験についての再調査や重要な共変量の存在を検討することを TSD2では提案されていた. このように特定の試験及び群について調査を行うきっかけとしても効力は有用と考えられる. 図 2-1 において、一般的に、c=3 のラインの外側のポイントは一般的にモデルによくあてはまっていないとみなされるため、何らかの説明が求められる可能性があるだろう.

	衣 2-2 回足効木モノル及い交重効木モノルによる p 延例来のアクノ アッシへの和木											
		固定	三効果モデル	V	変量効果モデル							
	meam SD median 95%信用区間				meam	SD	median	95%信用区間				
d12	-0.26	0.050	-0.26	(-0.36, -0.16)	-0.25	0.066	-0.25	(-0.38, -0.12)				
σ	-	-	-	-	0.11	0.055	0.10	(0.04, 0.25)				
$ar{\it D}_{res}$ †	46.8				41.9							
p_D	23.0				28.1							
DIC	69.8				70.0							

表 2-2 固定効果モデル及び変量効果モデルによる β 遮断薬のメタアナリシスの結果

10,000 回の稼働検査期間 (Burn-in) 後, 3 つの連鎖による 20,000 回の繰り返しサンプリングに基づく.

^{†44 (}データポイント) と比較

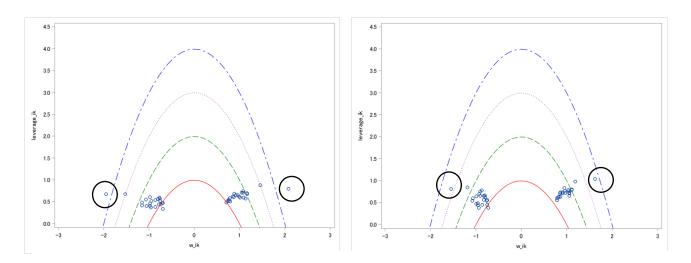


図 2-1 各データの残差逸脱度の平方根 $w_{ik}\left(\pm\sqrt{\overline{dev}_{ik}}\right)$ の効力($leverage_{ik}$)に対するプロット及び, $w_{ik}^2+leverage_{ik}=c$ (c=1 (実線), 2 (破線), 3 (ドット線), 4 (ドット-破線)) の曲線. (左図:固定効果モデル,右図:変量効果モデル)

2.4 一般化線形モデル

2.2 節で紹介した二項尤度に対する logit リンクのモデルを、様々なデータに適用できるモデルに拡張する. 本質的な考えは、メタアナリシスでの基本的な技法と同じだが、データの性質(連続値、割合(率)、カテゴリ)及びデータの発生プロセス(正規分布、ポアソン分布、多項分布など)を反映させるため、尤度とリンク関数を変更する. 一般化線形モデルの理論の枠組みでは、尤度を未知のパラメータγに関して定義し、このパラメータを±無限大の範囲に変換するためのリンク関数 g(·)を用いる. 式(2)に対するメタアナリシスのモデルは、

$$g(\gamma) = \theta_{ik} = \mu_i + \delta_{i,bk} I_{\{k \neq 1\}} \tag{5}$$

の形式による一般化線形モデルになる.ここに, θ_{ik} は通常,試験 i の治療群 k における治療効果の連続尺度(例えば対数オッズ)である. μ_i は試験 i における試験特有のベースライン効果で,局外母数 (nuisance parameter)として扱われる. $\delta_{i,bk}$ は試験 i の対照群 b に対する治療群 k の試験特有の治療効果である.例えば,試験 i の対照群 b (b=1) に対する治療群 2 の治療効果は

$$\delta_{i,12} \sim N(d_{12}, \sigma^2) \tag{6}$$

に従う.ここでは、2.2節と同様に、共通の分散に従うモデルを仮定している.

臨床試験では様々なタイプの応答のデータが得られることを考慮すると、一般化線形モデルによりそれらを解析することが求められる。メタアナリシスで基盤となるモデル(式(5)と(6))はそれぞれのケースで共通であり、尤度関数とリンク関数のみ異なる。ベイズ流の枠組みでは分散パラメータにおける事前分布の規定にも注意が必要である。表 2-3 に一般的に利用される尤度、リンク関数及び逆リンク関数の詳細を示す。

表 2-3 一般的なリンク関数及びその逆関数,並びに対応する尤度

リンクの名称	リンク関数	逆リンク関数	尤度関数	
サンクの名称	$\theta = g(\gamma)$	$\gamma = g^{-1}(\theta)$	儿 及)	
恒等式(identity)	γ	θ	正規	
Logit	$\ln \left(\frac{\gamma}{(1-\gamma)} \right)$	$\frac{\exp(\theta)}{1 + \exp(\theta)}$	二項,多項	
Log	ln(γ)	$exp(\theta)$	ポアソン	
Complementary log-log (cloglog)	$\ln\{-\ln(1-\gamma)\}$	$1 - \exp\{-\exp(\theta)\}$	二項,多項	
Reciprocal link	1/γ	$^{1}/_{\theta}$	ガンマ	
Probit	$\Phi^{-1}(\gamma)$	$\Phi(\theta)$	二項,多項	

また、**表 2-4** に、述べられているすべての様々な尤度に対して p_D を計算するために必要な残差逸脱度の公式とその予測式を示す.

モデルによる 残差逸脱度 尤度 予測式 Residual Deviance $\sum_{i} \sum_{k} 2 \left(r_{ik} \log \left(\frac{r_{ik}}{\hat{r}_{ik}} \right) + (n_{ik} - r_{ik}) \log \left(\frac{n_{ik} - r_{ik}}{n_{ik} - \hat{r}_{ik}} \right) \right)$ $r_{ik} \sim \text{Binomial}(p_{ik}, n_{ik})$ $\hat{r}_{ik} = n_{ik} p_{ik}$ $\sum_{i} \sum_{k} 2 \left((\hat{r}_{ik} - r_{ik}) + r_{ik} \log \left(\frac{r_{ik}}{\hat{r}_{ik}} \right) \right)$ $r_{ik} \sim \text{Poisson}(\lambda_{ik}E_{ik})$ $\hat{r}_{ik} = \lambda_{ik} E_{ik}$ $y_{ik} \sim N(\bar{y}_{ik}, se_{ik}^2)$ se_{ik}^2 は既知と仮定 \bar{y}_{ik} $\sum_{i} \sum_{k} 2 \left(\sum_{i} r_{ikj} \log \left(\frac{r_{ikj}}{\hat{r}_{ikj}} \right) \right)$ $r_{i,k,1:J} \sim \text{Multinomial}(p_{i,k,1:J}, n_{ik})$ $\hat{r}_{ikj} = n_{ik} p_{ikj}$ 多変量正規分布 $\sum_{i} (y_{i,1:k} - \bar{y}_{i,1:k})^{T} \Sigma^{-1} (y_{i,1:k} - \bar{y}_{i,1:k})$ $\bar{y}_{i,1:k}$ $y_{i,1:k} \sim N_k(\bar{y}_{i,1:k}, \Sigma_{(k \times k)})$

表 2-4 一般的な尤度に対する残差逸脱度及びモデルによる予測式

2.4.1 率データ:ポアソン尤度及び対数リンク

メタアナリシスに含まれる RCTs において、ある期間(試験ごとに異なり得る)での稀な事象の頻度 データが得られる場合、ポアソン尤度及び対数リンク関数の適用が考えられる。例えば、死亡数、また は医療機器が故障したイベント数などである。しかし、利用できるものは at risk の例数ではなく、at risk の観察人年 (person-years) の合計である。期間内に最終イベントに到達した患者の at risk の時間は試験 開始から最終イベントまでの時間になり、そうでない患者の at risk の時間は追跡期間と同じになる。な お、打ち切りも許容される。

 r_{ik} は追跡期間における試験 i, 治療群 k で生じたイベント数, E_{ik} は試験 i, 治療群 k での観察人年 (personyear), λ_{ik} は試験 i, 治療群 k での単位時間当たりのイベント発現率 (ハザード) であり, その尤度を

$$r_{ik} \sim \text{Poisson}(\lambda_{ik}E_{ik})$$

と示す. 関心のあるパラメータはハザード λ_{ik} , すなわち試験毎の各治療群で生じる単位時間当たりのイベントの発現率であり、通常、対数尺度でモデル化される. 等式(5)における線形予測子は対数尺度で、

$$\theta_{ik} = \log(\lambda_{ik}) = \mu_i + \delta_{i,bk} I_{\{k \neq 1\}} \tag{7}$$

となる. $\delta_{i,bk}$ は試験 i の対照群 b に対する治療群 k の試験特有の治療効果である. このモデルの重要な仮定は、各試験の各群で、追跡期間を通じてハザードが一定ということである. これは、すべての患者のハザードが同じである同質の集団のみにあてはまる. 追跡期間でハザードが一定であるが患者ごとにハザードが異なる集団では、ハザードの高い患者はイベントが早く起こりリスク集合から除外されていく傾向があるので、平均ハザードは時間に伴って減少していく.

このモデルは、再発事象のデータにも適用可能である。例えば、1人1回以上事故に遭遇する可能性がある事故の回数のデータでは、各群での事故の総数、すなわち事故の平均回数に患者数を掛け合わせたものに対しモデル化する。また、時間ではなく空間で繰り返し得られた観測値にもポアソンモデルが使用される。例えば、入れ歯の数である。繰り返しイベントが生じるデータにポアソンモデルをあてはめる場合は、イベントが独立であるという追加の仮定が必要となる。例えば、すでに1回事故に遭遇した人と、事故に1回も遭遇していない人の間で、事故に遭遇する確率は変わらない、とする仮定である。

2.4.2 率データ: 二項尤度及び cloglog リンク

あるメタアナリシスでは、試験毎に規定された追跡期間におけるイベントに到達する患者割合が報告されている。ただし、追跡期間は試験毎に異なる。追跡期間 f_i (日数及び週数などで測定)を伴う試験i,治療群kでのイベント数を r_{ik} で表すとき、等式(1)のように、各試験のデータ生成過程の尤度は二項分布になる。logit モデルを利用する場合、ある一定の追跡期間までにすべての患者がイベントを発生する、もしくは比例オッズの仮定(どの試験(どの区間)でもオッズが同じ)が保たれている、のいずれかの仮定が必要である。これらの仮定の下では、ハザード関数が複雑な形式になる。さらに追跡期間が長いほど多くのイベントが観察される場合は、標準的なlogit モデルは解釈が困難になる。試験ごとに追跡期間が異なる場合に適用できる最も単純な方法は、各試験の群毎にポアソン過程を仮定することである。それは試験i の群i でイベントが発生するまでの時間i i のボル、ハザードi i の指数分布

$$T_{ik} \sim Exp(\lambda_{ik})$$

に従うというものである. 試験iの群kにおける追跡期間 f_i までにイベントが発生しない確率、すなわち生存関数は、

$$Pr(T_{ik} > f_i) = exp(-\lambda_{ik}f_i)$$

となる. そのとき,各試験 i について,追跡期間 f_i の後に,試験 i の群 k におけるイベントを観測する確率 p_{ik} は,

$$p_{ik} = 1 - \Pr(T_{ik} > f_i) = 1 - \exp(-\lambda_{ik} f_i)$$
 (8)

と示される.これは時間に依存する形式になっている.ここでは,試験毎に異なる追跡時間 f_i を考慮したイベント発現率をモデル化する.等式(8)は対数ハザード $\log(\lambda_{ik})$ に対して非線形のため,リンク関数 complementary log-log (cloglog) は, $\log(\lambda_{ik})$ についての一般化線形モデルを得るために用いられ, $\theta_{ik} = \operatorname{cloglog}(p_{ik}) = \log(f_i) + \log(\lambda_{i,bk})$ を与える.そして $\log(\lambda_{ik})$ は等式(7)としてモデル化される.

$$\theta_{ik} = \operatorname{cloglog}(p_{ik}) = \log(f_i) + \log(\lambda_{i,bk}) = \log(f_i) + \mu_i + \delta_{i,bk} I_{\{k \neq 1\}}$$

ここに、治療効果 $\delta_{i,bk}$ は対数ハザード比を表す.

このモデルでは、追跡期間を通じてハザードが一定とみなす、ポアソンの割合モデルと同じ仮定を置いている。各試験でどの患者でもハザードが一定であることを意味しており、上記で述べたように強い仮定である。ただし、追跡期間の長さがイベント数に影響を与えないという仮定よりは好ましい。これらの仮定の臨床的な妥当性については、議論され、適切な文献、および解析に含まれる試験での追跡期間におけるハザードの変化を調査することにより、支持されるべきである。

ハザード一定の仮定が適切ではなく、追跡期間が長いほどイベント数が多いと考えられるとき、時間 依存的に変わるハザードを許容する方法がある. 1 つのアプローチは、区分定数ハザードの適用である. これらのモデルは、同一試験内で複数の追跡時間で報告されたデータがあるとき適用可能である.

別のアプローチは、追加の形状パラメータ α を含むワイブルモデルをあてはめることである.

$$Pr(T_{ik} > f_i) = \exp\{(-\lambda_{ik}f_i)^{\alpha}\}\$$

この場合,モデルは以下となる.

$$\theta_{ik} = \operatorname{cloglog}(p_{ik}) = \alpha(\log(f_i) + \mu_i + \delta_{i,bk}I_{\{k \neq 1\}})$$

非線形の予測子が用いられているため一般化線形モデルではなく,これらの拡張によりモデル化の自由度が大幅に高まった。しかしながら、より多くのデータを必要とすることを意味する。例えば、追加のワイブルパラメータは、追跡時間上の広い範囲にデータがあり、すべての治療に対して同じ形状パラメータを仮定できる場合にのみ適切に特定される。

2.4.3 競合リスク:多項尤度及び対数リンク

複数の相互に排他なエンドポイントが定義されており、患者がそのいずれかに達するとリスク集団から外れる場合、競合リスク解析の適用が考えられる。例えば、統合失調症の治療薬の試験で、再発、許容できない副作用による治療中止、他の理由による治療中止のいずれかが生じるまで観測を継続する場合である。試験終了まで状態を維持している患者は打ち切りとなる。この競合アウトカムの間の統計的依存関係はモデルの中で考慮される必要がある。この依存関係は実質的には試験内のもので、アウトカム間の負の相関として試験毎の群別に適用される。アウトカムイベントの発現は確率過程であるため、

もし多くの患者がいずれか 1 つのアウトカムに到達する場合,その他のアウトカムにはほとんど到達しないはずである.複数の試験から,試験 i の追跡期間 f_i の最後で,互いに排他なエンドポイント j=1,2,...,J のそれぞれに到達する試験 i の群 k の患者数 r_{ikj} が報告されている.この場合,応答 r_{ikj} は以下の多項分布に従う. 関心のあるパラメータは,患者が最初の状態からエンドポイント j のいずれかに移動するハザード λ_{iki} である.

$$r_{i,k,j=1,...,l} \sim \text{Multinomial}(p_{i,k,j=1,...,l}, n_{ik}), \quad \sum_{i=1}^{J} p_{i,k,j} = 1$$
 (9)

J番目のエンドポイントが打ち切りの観測値を表している,すなわち,患者が追跡期間中に他のエンドポイントのいずれにも到達しない場合である.観察期間 f_i において一定のハザード λ_{ikj} を仮定できる場合,試験 i における群 k について,アウトカム j が観測期間中に生じる確率は,

$$p_{ikj}(f_i) = \frac{\lambda_{ikj}}{\sum_{u=1}^{J} \lambda_{iku}} \left[1 - \exp(-f_i \sum_{u=1}^{J-1} \lambda_{iku}) \right], j = 1, 2, 3, \dots, J - 1$$

になる. 最初の状態を維持している確率, すなわち打ち切りになる確率は, 単純に 1 から J-1 の注目している状態のいずれかに到達している確率の合計を引いて,

$$p_{ikJ}(f_i) = 1 - \sum_{u=1}^{J-1} p_{iku}(f_i)$$

となる. 関心のあるパラメータはハザード λ_{iki} で、これらは対数尺度でモデル化される.

$$\theta_{ikj} = \log(\lambda_{ikj}) = \mu_{ij} + \delta_{i,bk,j}I_{\{k\neq 1\}}$$

ここに、 $\delta_{i,bk,j}$ は試験iでのアウトカムjにおける対照群bに対する群kの試験特有の効果であり、試験iでのアウトカムjにおける対照群1に対する群2の試験特有の治療効果は、

$$\delta_{i,12,j} \sim N \big(d_{12j}, \sigma_j^2 \big)$$

である. 変量効果分布の試験間分散 σ_j^2 は,各アウトカムjで特有である. 分散については以下の 3 つのモデルが考えられる. それらは,①試験間分散を $\sigma_j^2 = 0$ とした固定効果モデル,②試験間分散を $\sigma_j^2 = \sigma^2$ とした Random Effects Single Variance モデル(試験間分散を各アウトカムに対して同じと想定),及び ③各アウトカムjに対して異なる試験間分散 σ_j^2 をもつ Random Effects Different Variances モデル(試験間分散を各アウトカムに対して同じではないと想定)である.

競合リスクモデルは、各アウトカムで時間に依らずハザードが一定である、すなわち比例ハザード性を仮定しており、これは前述の cloglog モデルでの仮定と同じである。更に、各アウトカムのリスク比が期間を通じて一定である(比例競合リスク)という追加の仮定をおいている。この仮定を緩めるさらなる拡張の方法も検討されている。

2.5 間接比較への拡張とネットワークメタアナリシス

ここまでの定義から,変量効果モデルにおけるメタアナリシスでは,M個の試験それぞれによる治療1に対する治療2の効果 $\delta_{i,12}$ は,平均 d_{12} ,分散 σ_{12}^2 の正規分布に従い,試験間で交換可能と仮定される.ここでは,同様の一連の試験(すなわち,同様の研究課題を伴う試験であり,患者対象集団,治療法,エンドポイントなどが類似している)で治療1と治療3の比較を行う場合も想定する.治療1と治療3の2群間での変量効果モデルでのメタアナリシスを実行するため群間差 $\delta_{i,13}$ は, $\delta_{i,13} \sim N(d_{13}, \sigma_{13}^2)$ に従って交換可能であると仮定する.この時,治療2と治療3の試験毎の治療効果の比較 $\delta_{i,23}$ もまた交換可能であると仮定する.

$$\delta_{i,23} \sim N(d_{23}, \sigma_{23}^2)$$

これは移行性(Transitivity)の関係, すなわち $\delta_{i,23} = \delta_{i,13} - \delta_{i,12}$ から導かれる. したがって,

$$d_{23} = d_{13} - d_{12} \tag{10}$$

および,

$$\sigma_{23}^2 = \sigma_{12}^2 + \sigma_{13}^2 - 2\rho_{23}^{(1)}\sigma_{12}\sigma_{13}$$

である.ここに $\rho_{23}^{(1)}$ は,試験内における治療 1 に対する治療 2 の相対治療効果と,治療 1 に対する治療 3 の相対治療効果の間の相関を表す.単純化のために,分散は等分散 $\sigma_{12}^2 = \sigma_{13}^2 = \sigma_{23}^2 = \sigma^2$ であると仮定している.これは多群試験における 2 治療間での相関を 0.5 と想定している場合に相当する(補足: σ_{23}^2 の式に等分散の式を代入).異質性を許容する分散モデルについては Lu & Ades [4]を参照されたい.

治療効果 $\delta_{i,12}$ と $\delta_{i,13}$ に関する交換可能性の仮定により、治療 1 と治療 2 の比較試験、治療 1 と治療 3 の比較試験から、治療 2 と治療 3 の間接比較が可能になる。また治療 1 と治療 2 を含む試験、治療 1 と治療 3 を含む試験の統合では、治療 2 と治療 3 の試験を含むことも許容する。2 群のメタアナリシスの標準的な仮定と、間接比較及び、直接比較と間接比較を含む混合比較(Mixed treatment comparison; MTC)で求められる仮定との関係性に注意が必要となる。すなわち、変量効果の 2 群メタアナリシスでは、治療 1 と治療 2 を比較した試験で、効果 $\delta_{i,12}$ に交換可能性の仮定を置き、治療 1 と治療 3 を比較した試験では、効果 $\delta_{i,13}$ に対して交換可能性の仮定が必要である。NMA では、この治療 1 と治療 2 との比較、及び治療 1 と治療 3 との比較の両方の治療効果に対して、交換可能性の仮定を置かなければならない。この仮定は、追加の治療 k=4,5,...,s にも拡張され、全ての治療効果の交換可能性を仮定する必要がある。さらに一致性の等式(consistency equations)

$$\begin{split} d_{23} &= d_{13} - d_{12} \\ d_{24} &= d_{14} - d_{12} \\ &\vdots d_{(s-1),s} = d_{1s} - d_{1,(s-1)} \end{split}$$

も仮定される(ネットワークでの直接比較と間接比較の結果が一致していること). これらは間接比

較及び混合比較のために必要な仮定で,全ての試験が同様の研究課題であれば,これらの仮定は自然である.しかしながら,理論上,治療効果の一致性が担保される必要がある一方で,実際に得られる結果では直接比較と間接比較の結果が一致しない可能性がある.一致性の等式は,(s-1)個の基本パラメータ d_{12} , d_{13} ,…, d_{1s} で構成されており,各基本パラメータには事前分布が設定される.一致性の等式により,基本パラメータ以外の相対効果は,基本パラメータの関数で表される.なお,各試験で対照として設定されている治療を対照治療又は対照群(本書ではb群として参照)と呼び,ネットワークメタアナリシスにおいて基準とする治療を参照治療又は参照群(本書では第1群として参照)と呼ぶ.ただし,対照群と参照群が一致する場合は対照群を第1群として参照している.

複数の治療が比較される場合を想定して表記を改める. 対照治療(群1)に対して, 群 kの治療における試験特有の治療効果は, 以下の共通の変量効果の分布から抽出される.

$$\delta_{i,1k}{\sim}N(d_{t_{i1},t_{ik}},\sigma^2)$$

ここに、 $d_{t_{i1},t_{ik}}$ は、試験 iの群 1 における治療 t_{i1} に対する試験 iの群 kにおける治療 t_{ik} の平均効果を表す。治療 1 と治療 2 を比較する試験では $d_{t_{i1},t_{ik}}=d_{12}$ 、治療 2 と治療 3 を比較する試験では $d_{t_{i1},t_{ik}}=d_{23}$ である。治療 2 に対する治療 3 の(統合)治療効果 d_{23} は式(10)から推定される。

2.6 多群試験の組み込み

ここでは関心のある治療法を複数含む多群試験を想定する. 多群試験のメタアナリシスの実施方法については、Lu & Ades[5]によるベイズ流の枠組で、そして Lumley[6]と Chootrakool & Shi[7]による頻度論の枠組で検討されている.

試験間の交換可能性の仮定に基づき,単一の多群試験では変量効果のベクトル δ_i を推定する.例えば、3 群試験では 2 つの変量効果(治療 1 に対する治療 2,3 の治療効果),4 群試験では 3 つの変量効果(治療 1 に対する治療 2~4 の治療効果)を生成する. これまでと同様,全ての相対治療効果は同じ試験間の分散をもち、

$$\boldsymbol{\delta_{i}} = \begin{pmatrix} \delta_{i,12} \\ \vdots \\ \delta_{i,1a_{i}} \end{pmatrix} \sim N_{a_{i-1}} \begin{pmatrix} d_{t_{i1},t_{i2}} \\ \vdots \\ d_{t_{i1},t_{ia_{i}}} \end{pmatrix}, \begin{pmatrix} \sigma^{2} & \sigma^{2}/2 & \cdots & \sigma^{2}/2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{2}/2 & \sigma^{2}/2 & \cdots & \sigma^{2} \end{pmatrix}$$

$$(11)$$

に従うとする. ここに、 δ_i は変量効果のベクトルで、多変量正規分布に従い、試験 i での群の個数を a_i ($a_i=2,3,...$)、そして $d_{t_{i1},t_{ik}}=d_{1,t_{ik}}-d_{1,t_{i1}}$ を示す。治療群 2 からk-1までの結果が全て所与の下、群 k>2の変量効果について、その条件付き単一変量分布は、

$$\delta_{i,1k} | \begin{pmatrix} \delta_{i,12} \\ \vdots \\ \delta_{i,1(k-1)} \end{pmatrix} \sim N \left(\left(d_{1,t_{ik}} - d_{1,t_{i1}} \right) + \frac{1}{k-1} \sum_{j=1}^{k-1} \left[\delta_{i,1j} - \left(d_{1,t_{ij}} - d_{1,t_{i1}} \right) \right], \frac{k}{2(k-1)} \sigma^2 \right)$$
(12)

である. 各多群試験に対する変量効果を推定するためには、式(11)での多変量分布と式(12)での条件付き分布のどちらかが利用され、パラメータ間の相関が考慮される. 式(12)の条件付き分布の利用は群の個数に寄らず対応できるため、より一般的なものになっている. これは、群のいくつかが無作為に欠測(Missing At Random; MAR) しているが、S 個の治療を含む M 個の試験の連結されたネットワークとして検討できる. (MAR は、群の選択がランダムであるということを意味しているわけではなく、群の欠落がその治療の有効性に無関係ということを意味している.)

2.7 正規尤度及び恒等式のリンク関数を伴うモデルの連続量データへの適用

応答変数が連続量の場合、メタアナリシスは標本平均 y_{ik} とその標準誤差 se_{ik} に基づく、サンプルサイズが極端に小さくない限り、例え個々のデータ分布が歪んでいても、標本平均は中心極限定理により近似的に正規分布に従い、尤度は以下で示すことができる。

$$y_{ik} \sim N(\theta_{ik}, se_{ik}^2)$$

ここに関心のあるパラメータは平均 θ_{ik} である. リンク関数として恒等式 (identity) が利用され、線形モデルは自然尺度上で

$$\theta_{ik} = \mu_i + \delta_{i,bk} I_{\{k \neq 1\}} \tag{13}$$

と表すことができる.ここに, $I_{\{u\}}=\begin{cases} 1 & \text{if } u \text{ is true} \\ 0 & \text{otherwise} \end{cases}$ で, μ_i は試験 i における試験特有のベースライン 効果, $\delta_{i,bk}$ は試験 i における対照治療(b)に対する治療 k の効果を示す.なお,固定効果モデルを適用する場合は $\delta_{i,1k}=d_{1k}$,変量効果モデルを適用する場合は $\delta_{i,1k}\sim N(d_{1k},\sigma^2)$ となる.

表 2-5 のデータはパーキンソン病における補助療法としてドーパミンアゴニストが与えられた患者における平均オフ時間(薬剤の効果が認められない時間)を示している. 利用できるデータは, 5 つの薬剤(1(プラセボ)と 2~5 の治療薬)の 7 試験における各群の平均,標準偏差,例数である. そのネットワーク図を図 2-2 に示す.

表 2-5 パーキンソン病の事例:治療1をプラセボ,治療2-5を治療薬としたときの,試験名,治療法,各群の例数,標準偏差(SD)を伴った平均オフ時間の減少,治療群間差とその標準誤差(SE)

試験	治療法	例数	平均オフ時間の減少	SD	群間差	SE (群間差)
1	1	54	-1.22	3.7		
1	3	95	-1.53	4.28	-0.31	0.668
2	1	172	-0.7	3.7		
2	2	173	-2.4	3.4	-1.7	0.383
	1	76	-0.3	4.4		
3	2	71	-2.6	4.3	-2.3	0.718
	4	81	-1.2	4.3	-0.9	0.695
4	3	128	-0.24	3		
4	4	72	-0.59	3	-0.35	0.442
5	3	80	-0.73	3		
3	4	46	-0.18	3	0.55	0.555
6	4	137	-2.2	2.31		
0	5	131	-2.5	2.18	-0.3	0.274
7	4	154	-1.8	2.48		
'	5	143	-2.1	2.99	-0.3	0.32

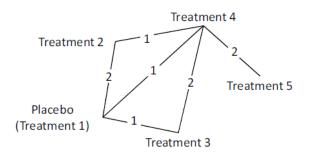


図 2-2 パーキンソンのネットワーク:ノード(節点) は治療法(治療コード)を表し、それらを結ぶ実線[辺 (edge)]はRCTで直接比較された治療の対を示し、その実線上の数値は試験数である. (再掲載)

[出典: NICE DSU TSD2 (http://nicedsu.org.uk/technical-support-documents/technical-support-documents/)]

表 2-5 の治療群単位のデータを用いたモデルで実装した WinBUGS プログラム(データ,固定効果モデルと初期値,変量効果モデルと初期値)と,その結果として相対治療効果の事後分布の要約を示す.本稿では,固定効果モデルと変量効果モデルで異なる部分は Bold で示し,モデルのプログラムについては,行ごとに注釈を付している.なお,TSD2 では同じパーキンソン病のデータに対して,群間差のデータを用いたモデル,及び治療群単位のデータと群間差のデータの両方が混在するモデル(shared parameter model)についても紹介されている.

<データ>

最大の治療群数は3である. そのため,治療の指標として3つのベクトルt[,1],t[,2],t[,3]. 連続量での結果y[,]とそれらの標準誤差se[,],及び群数na[]が必要である.

```
# Data (Parkinson's example)
list(ns=7, nt=5)
t[,1]
          t[,2]
                 t[,3]
                        y[,1]
                               y[,2]
                                       y[,3]
                                              se[,1]
                                                     se[,2]
                                                            se[,3]
1 3
          NA
                 -1.22
                        -1.53
                               NA
                                       0.504
                                              0.439
                                                     NA
1
  2
          NA
                 -0.7
                        -2.4
                                NA
                                       0.282
                                              0.258
                                                     NA
                                                             2
1
  2
          4
                 -0.3
                        -2.6
                               -1.2
                                       0.505
                                              0.510
                                                     0.478
                                                            3
3 4
          NA
                 -0.24
                       -0.59
                               NA
                                       0.265
                                              0.354
                                                     NA
                                                             2
3 4
          NA
                 -0.73
                       -0.18
                               NA
                                       0.335
                                              0.442
                                                     NA
                                                             2
4 5
          NA
                 -2.2
                        -2.5
                               NA
                                       0.197
                                              0.190
                                                     NA
                                                             2
4 5
          NA
                 -1.8
                        -2.1
                                       0.200 0.250
                               NA
                                                     NA
                                                             2
```

<固定効果モデル>

変量効果モデルではコードの変更が必要な個所を Bold で示す.

```
# 多群試験への固定効果モデル
model{
 for(i in 1:ns){
                                                # 試験の繰り返し
   mu[i] \sim dnorm(0,.0001)
                                                # 全ての試験のベースラインに曖昧な事前分布を設定する
   for (k in 1:na[i]) {
                                                # 治療群の繰り返し
       var[i,k] \leftarrow pow(se[i,k],2)
                                                # 分散を計算
                                               # 精度に変換する
       prec[i,k] <- 1/var[i,k]
       y[i,k] ~ dnorm(theta[i,k],prec[i,k])
                                               # 正規尤度
                                               # 線形予測子のモデル
       theta[i,k] <- mu[i] + d[t[i,k]] - d[t[i,1]]
                                                       # 逸脱度への寄与
       dev[i,k] \leftarrow (y[i,k]-theta[i,k])*(y[i,k]-theta[i,k])*prec[i,k]
                                                #この試験における残差の逸脱度の寄与の合計}
   resdev[i] <- sum(dev[i,1:na[i]])
 totresdev <- sum(resdev[])
                                                # 残差の逸脱度の合計
                                                # 参照治療の治療効果をゼロに設定
 for (k \text{ in } 2:nt) \{ d[k] \sim dnorm(0,.0001) \}
                                                # 治療効果の無情報事前分布
```

<固定効果モデルに対応する初期値(3通り)>

```
# chain 1
list(d=c( NA, 0, 0, 0, 0), mu=c(0, 0, 0, 0, 0, 0))
# chain 2
list(d=c( NA,-1,-3,-1, 1), mu=c(-3,-3,-3,-3,-3,-3))
# chain 3
list(d=c( NA, 2, 2, 2, 2), mu=c(-3, 5,-1,-3, 7,-3,-4))
```

<変量効果モデル>

固定効果モデルからコードの追記及び変更となる個所を Bold で示す.

```
# 多群試験への変量効果モデル
model{
  for(i in 1:ns){
                                    # 試験の繰り返し
   w[i,1] < 0
                                    # コントロール群に対する多群試験の調整を0とする
   delta[i,1] <- 0
                                    # コントロール群の治療効果をゼロとする
   mu[i] \sim dnorm(0,.0001)
                                   # 全ての試験のベースラインに無情報事前分布を設定する
   for (k in 1:na[i]) {
                                   # 治療群の繰り返し
     var[i,k] \leftarrow pow(se[i,k],2)
                                   # 分散を計算
     prec[i,k] <- 1/var[i,k]
                                   # 精度に変換する
     y[i,k] ~ dnorm(theta[i,k],prec[i,k])
                                   # 正規尤度
```

```
# 線形予測子のモデル
     theta[i,k] <- mu[i] + delta[i,k]
     dev[i,k] <- (y[i,k]-theta[i,k])*(y[i,k]-theta[i,k])*prec[i,k] # 逸脱度への寄与
   resdev[i] <- sum(dev[i,1:na[i]])
                                     # この試験における残差の逸脱度の寄与の合計
   for (k in 2:na[i]) {
                                     # 治療群の繰り返し
     delta[i,k] ~ dnorm(md[i,k],taud[i,k]) # 試験特有の LOR 分布
     <u>md[i,k] <- d[t[i,k]] - d[t[i,1]] + sw[i,k]</u> # 治療効果の分布の平均(多群試験の相関を伴う)
     taud[i,k] <- tau *2*(k-1)/k
                                      # 治療効果の分布の精度(多群試験の相関を伴う)
     w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]]) # 多群 RCTs に対する調整
     sw[i,k] <- sum(w[i,1:k-1])/(k-1)
                                     # 多群試験に対する累積の調整
   }
 }
totresdev <- sum(resdev[])
                                      # 残差の逸脱度の合計
                                      # 参照治療の治療効果をゼロに設定
for (k \text{ in } 2:nt) \{ d[k] \sim dnorm(0,.0001) \}
                                     # 治療効果に無情報事前分布を設定
sd ~ dunif(0.5)
                                      # 試験間 SD に無情報事前分布を設定
tau <- pow(sd,-2)
                                      # 試験間の精度 = (1/試験間分散)
```

<変量効果モデルに対応する初期値(3通り)>

表	2-6 19-	キンソン	/の事例	•	相対治療効果の事後分布の要約	1

		固定	三効果モデ	ジル	変量効果モデル			
	mean	SD	median	95%信用区間	mean	SD	median	95%信用区間
d_{12}	-1.81	0.33	-1.81	(-2.46,-1.16)	-1.85	0.54	-1.84	(-2.91,-0.85)
d_{13}	-0.47 0.49 -0.47		(-1.43,0.49)	-0.5	0.66	-0.5	(-1.78,0.75)	
d_{14}	-0.52	0.48	-0.52	(-1.46,0.43)	-0.53	0.65	-0.53	(-1.77,0.71)
d ₁₅	-0.82	0.52	-0.82	(-1.84,0.22)	-0.83	0.8	-0.83	(-2.35,0.69)
σ	-		-	-	0.4	0.43	0.28	(0.01,1.55)
$\overline{\it D}_{res}$ †	13.3				13.6			
pD	11				12.4			
DIC	24.3				26			

 d_{1k} : 治療法 k の治療法 1(プラセボ)に対する相対治療効果、稼働検査期間 (Burn-in) 50,000、サンプリング回数 100,000 (3-chain)、いずれの相対治療効果の事前分布も N(0,1002) † 15 (データポイント) と比較

表 2-6 から, 固定効果モデルと変量効果モデルで, 残差逸脱度や DIC は同程度であるとみなされた. 両モデルとも治療法 2 の治療法 1(プラセボ)に対するオフ時間減少の 95%信用区間が 0 を含まないため, 治療法 2 はプラセボと比べオフ時間を減少させることが示唆された.

外部ソースから次のコードを加筆し、治療1における結果である平均-0.73、精度 (分散の逆数)21の値

 $A \sim dnorm(-.73,21)$ for (k in 1:nt) { T[k] <- A + d[k] }

2.8 ベイズ流 MCMC 法における技術的な課題

WinBUGS のベイズ流 MCMC パッケージの利用は便利であるが、注意も必要である. 以下に TSD2 に 記載されていた注意をまとめる、収束の確認においては大きく異なる初期値を利用して少なくとも3つ の連鎖を実行し、収束後、各連鎖が同じ事後分布から抽出されていること確認することが提案されてい る. 最初の稼働検査期間(以下, Burn-in)と事後分布の標本数は保守的に大き目に設定すべきで,これ らの繰り返し数も解析結果と共に報告されるべきである. シミュレーションの回数と自己相関の程度を 反映しているモンテカルロエラーが、関心のあるパラメータの事後分布の標準偏差の 5%未満であるべ きと提案されているものもある[8]. また、ベイズ流の MCMC に適用するエビデンス統合のモデルにお いて, 参照治療の選択, 事前分布の選択, 二項分布とポアソン分布におけるゼロセルカウント(収束性) など,注意深い検討が必要となる.参照治療の選択により事後分布の推定値に影響を与える可能性があ るが、各パラメータの事前分布に無情報事前分布を設定することでその影響を無視できる、そのため、 参照治療はプラセボや標準治療といった解釈が容易なものを設定すべきである. また, イベント数(例 えば、有害事象の発現例数)が 0 (例) となる零セルを伴う二項尤度とポアソン尤度が許容されるのは ベイズ流の MCMC アプローチの主要な特徴の一つである. 対数オッズ比や対数相対リスクに対する頻 度論のアプローチでは通常ゼロセルに 0.5 など任意の定数を加える必要があるが、ベイズ流 MCMC ア プローチではゼロセルを伴う試験のためのその様な操作は必要ない[ただし,スパースデータに対して は収束性やモデルの安定性に関して別途検討が必要となる場合もある(詳細は TSD2 を参照)]. また, 得られる事後分布の標本はパラメータ間の相関が保持され、確率論的な費用効果分析にそのまま用いる ことができるという利点もある.

参考文献:

- [1] Sutton, A.J., Higgins, J. Recent developments in meta analysis. Statistics in Medicine 2008; 27(5):625-650.
- [2] Dias, S., Ades A.E., Welton N.J., et al. Network Meta-Analysis for Decision-Making. John Wiley & Sons, Inc. 2018.
- [3] Welton N.J, Sutton A.J., Cooper N., Abrams K.R., Ades A.E.. Evidence Synthesis for Decision Making in Healthcare. John Wiley & Sons, Inc. 2012.
- [4] Lu, G., Ades, A. Modeling between-trial variance structure in mixed treatment comparisons. Biostatistics 2009; 10(4):792-805.
- [5] Lu, G., Ades, A.E. Combination of direct and indirect evidence in mixed treatment comparisons. Statistics in Medicine 2004; 23(20):3105-3124.
- [6] Lumley, T. Network meta analysis for indirect treatment comparisons. Statistics in Medicine 2002; 21(16):2313-2324.
- [7] Chootrakool, H., Shi, J.Q. Meta-analysis of multi-arm trials using empirical logistic transform. The Open Medical Informatics Journal 2008; 2:112-116.
- [8] WinBUGS User Manual Version 1.4, January 2003 [cited June 24, 2014]. Avairable from: http://www.mrc-bsu.cam.ac.uk/wp-content/uploads/manual14.pdf.

3 異質性:サブグループ、メタ回帰、バイアス及びその補正(TSD3)

TSD3 では、治療法の相対治療効果の異質性に焦点を当てている。異質性は効果修飾因子の存在を示しており、患者集団又は設定が異なる状況での治療効果の真のバラツキと、試験の実施方法に関連する偏り(以下、バイアス)とを区別する。治療の相対治療効果のバラツキ(異質性)は、エビデンスの外部妥当性を脅かし、すなわち、結果の一般化可能性を制限するものであり、一方、臨床試験の実施上の不完全性は内部妥当性への脅威となる。いずれの場合も、共変量の値により患者を無作為化できないため(補足:試験内では動的割付又は層別割付などにより達成可能だが、試験間で無作為化はできない)、効果を修飾する共変量に関する研究は、あらゆる点で観察研究の一形態となることに注意が必要である。本TSDは、2群比較のメタアナリシス、間接比較、及びネットワークメタアナリシス(NMA)における外れ値の検出、メタ回帰、及びバイアス補正の手法に関する事例を用いたガイダンスとなっている。特に費用効果分析での異質性の意味合いに関する明確なガイダンスを与えている。また平均効果の分布よりも、「新しい」試験における治療効果の予測分布は、費用対効果評価の意思決定において、多くの場合より適切とされている(補足:後述するように、費用対効果評価で取り扱う効果の異質性(バラツキ)は、全体の効果の平均値が取り得る範囲ではなく、個々の試験が取り得る範囲を考慮する必要があるとされ、そのため予測分布を使用することが推奨されている)。さらに異質性を検討する際には、バイアスによる真の変動と、ランダムなバラツキを考慮する必要があるとされている。

サブグループごとに治療効果が異なることが疑われる場合でも、交互作用項を含む単一モデルでの解析は、各サブグループに対し別々の交互作用モデルを考える解析より優れていることが示唆される. 試験レベルで効果を修飾する共変量が存在する可能性がある場合、3つのタイプのメタ回帰モデルを、ネットワークメタアナリシス(NMA)のフレームワークにて検討可能である. 共変量と治療法(各治療法 vs 参照治療)との交互作用に関して、以下の3つを検討する.

- (1) 各治療法に対し(関連しない)別々の交互作用項[各治療法の交互作用は,互いに影響与え合わない]
- (2) 交換可能で(関連する)交互作用項[各治療法の交互作用は同一の確率分布に従う]
- (3) 全ての治療法で、単一で共通の交互作用項[全ての治療法の交互作用は同一]

単一の交互作用項が、意思決定において最も有用である可能性が高いものと考える(補足:この記述は普遍的なものではなく、費用効果分析で使用する場合のNICEの見解である). 連続値の共変量に対するベイズ流のメタ回帰、及び「ベースライン」リスクに対するメタ回帰に関して、理解を助ける事例を示し、その結果の解釈が与えられている. 個々の患者データ (IPD) を用いたメタ回帰では、共変量の分布がはるかに広いため、効果修飾因子をより高い精度で推定できる. また、いくつかの試験からは IPD、他のいくつかの試験からは要約データを組合せる方法の概略も説明されている. さらに各事例に対するWinBUGS コードも提供されている. 各メタ回帰プログラムの大部分は、TSD2[8]のプログラムと同じである. したがって、プログラムコードのモジュール性を強調するために、主な違いが bold で強調されて

いる. 固定効果モデルのプログラムコードは、変量効果モデルと同じデータ構造を用いて実行可能である. また、プログラム 2~6 のコードは一般的なものであり、任意の数の治療群及び複数治療群の試験での2群のメタアナリシス、又はNMAに適用できる. また、様々なプログラムに対し、推定結果をモニターし、モデルチェックする関連パラメータの計算法を示した. 異質性パラメータがないことを除いて、固定効果モデルにてモニターするべきパラメータは、変量効果モデルのパラメータと同じである.

なお、TSD でも推奨されているが、WinBUGS プログラムは TSD 発行以降に修正されている場合があり、以下の Web サイトから最新プログラムをダウンロードすることを強く推奨する.

ダウンロード: http://nicedsu.org.uk/technical-support-documents/evidence-synthesis-tsd-series/

最後の3.8 節で、バイアス補正のための方法の概要を説明する。それらは、①メタ回帰、②研究の質の低さの指標に関連したバイアスを補正するため、外部データに基づいた事前分布を利用する方法、③ネットワーク内に含まれる研究の質に関連したバイアスを推定及び補正するため、ネットワークによるエビデンス統合の利用、及び④専門家の意見を反映したバイアスの事前分布の利用の4つである。

3.1 概要

TSD3 は異質性に関する文書であり、具体的には相対治療効果の試験間のバラツキに関係している. NICE ガイドライン・テクニカル・アプレイザル[1]で推奨されているように、異質性の理由を調べるために使用できる手法ガイダンスの提供を目的としている.「ベースライン」の自然病歴(natural history)のバラツキは TSD5[2]で扱われている。本シリーズの他のドキュメントと共通して、意思決定での様々な異質性の影響、異質性の潜在的な原因を推定又は調整できるモデルの技術仕様、及び意思決定の際のモデルの解釈に特に焦点を当てている。異質性の起源と含意(影響)に関してかなりの文献があり、その問題の導入及びさらなる参考文献に関してはコクランハンドブック[3]を参照されたい.

治療効果の異質性は、効果修飾メカニズムの存在、すなわち治療効果と試験との交互作用、又は治療効果と試験レベルの変数(男性の割合、1 次予防の患者のみの試験など)との交互作用の存在を示している。通常、2 種類の交互作用の違いが区別される。まず後者の交互作用(すなわち、治療効果と試験レベルの交互作用)は、試験間での異なる患者集団、設定、又はプロトコルのバリエーションに起因する治療効果のバラツキから生じる(治療効果の臨床的バラツキと呼ぶ)。このバラツキは、試験の外的妥当性に対する脅威であり、試験結果を一般化できる程度を制限している。臨床試験は、特定の状況での治療効果の不偏推定値を提供し得るが、特定の決定問題での対象集団に関しては「偏っている」かもしれない。選択基準と除外基準を慎重に考慮することで、この種のバイアスを最小限に抑えることができる(補足:HTA に含める薬剤等の臨床試験を自社で実施する場合は、自社の臨床試験の組入除外基準を指し、そうでない臨床試験に関してはシステマティックレビューでの各試験の選択除外基準を考慮する)。次に前者の交互作用(すなわち、治療効果と試験の交互作用)は、内的妥当性を脅かすものであり、臨床試験の実施方法の欠陥に起因する。そのような試験では、意思決定のための対象集団と同じであるかどうかに関わらず、治療効果の偏った推定値を与える。これらのバイアスは、臨床試験全体を通じて、その大きさが確率的に(ランダムに)変化すると考えられ、必ずしも平均がゼロとは限らない(補足:

バイアスがなければ、その平均値はゼロに近づく). 最も明確な例は、割付の密封 (隠すこと) の欠如 (補足: 盲検性が成立していない)、又は二重盲検性の欠如 (補足: 割付表が開封されていなくても、ある検査値を測定すると医師には判ってしまうケース) など、試験の質が低いことによるバイアスであり、これらはより大きな治療効果と関連している. [4][5] 両方の交互作用を包含した異質性のモデルは、Higgins et al.[6]により提案されている. 異質性の原因が何であるか、又は臨床的な因子の真の変動によるものがどの程度なのか、他の未知の原因によるものがどの程度なのかを、決定できることはほとんどない.

TSD3 では、異質性を取扱うメタ回帰及びバイアスの調整方法のガイダンスを提供する. NMA では、治療の相対治療効果のバラツキもまた、2 群比較での不一致性(TSD4[7]参照)を引き起こす可能性がある. したがって、ここで紹介する方法は、不一致性の取扱いにも適している. 特に断りのない限り、異質性を検討する場合、異質性及び不一致性(の両方)を検討することと解釈できる.

本文書は、NMA、間接比較及び2群比較のメタアナリシスへの一般化線形モデリングの枠組みを設定する TSD2[8]を補助する文書である. TSD2[8]では、どのように同じコアモデルを、様々な尤度及びリンク関数に適用可能かを説明している. これは以下に示す交差検証法(以下、クロスバリデーション;3.6節)、サブグループ解析又はメタ回帰(3.7節)及びバイアスの調整(3.8節)で検討するベイズモデルに、完全に引き継がれることを理解されたい.

3.2 メタ回帰の概要

メタ回帰は、観察された試験間の異質性の一部又は全部を説明するために、メタアナリシスから得られた治療効果の大きさと、解析に含まれた臨床試験の特定の数的な特性との関係を検討するのに使用される。これらの特性とは、試験に参加した個々の患者の特徴に起因する可能性があり、又は臨床試験の設定や実施に直接起因する可能性がある。他の形態のメタアナリシスと同様に、メタ回帰は、(試験レベルで)要約されたアウトカム及び(要約された)共変量に基づいたり、又は個別患者データ (IPD)が利用可能であり得る場合はそれに基づいたりする。教科書[3][9]では、無作為化比較試験 (RCT) データに関心を限定したとしても、効果修飾因子の検討は本質的に観察的であることを強調している。これは患者を共変量の1つの値又は他の値に無作為化することが不可能であるためである。その結果、本文書で説明するメタ回帰は、無作為化されない研究に付随する解釈及び推論のすべての困難を継承する。(困難とは)交絡、共変量間の相関、そして最も重要なことは、関連性からは因果関係を推論不可能であることである。しかし、メタ回帰に基づく推論での信頼性に対して、上記の制約が広く適用されるが、以下で説明するように、関心のある共変量及びデータの構造の性質に依存して、メタ回帰のエビデンスの質に大きな違いが生じる。

3.2.1 試験内及び試験間の共変量

試験レベルの共変量は、試験特性に関連する共変量、あるいは IPD 又は患者背景で分割された適切な結果が利用できない場合、試験レベルで集計された患者背景に関連する共変量として定義する。患者レベルの共変量は、患者属性に関連する共変量として定義される。また、患者レベルの共変量は、IPD が

利用可能であるか、又は患者背景で十分に分割された結果(アウトカム)が提供されている場合、患者属性に関連し、各試験の特定の患者に帰することができる共変量として定義される.

まずは、カテゴリの共変量に関して、シナリオを以下のように区別する.

A1. 試験特性に関連する試験レベル共変量

例えば、1次予防及び2次予防患者集団で実施された試験.この共変量は、試験間の治療と共変量の 交互作用に関連する.解析方法は、サブグループ効果 (3.7.1 節) を参照のこと.

A2. 患者特性に関連する試験レベル共変量

以下の事例が含まれる.

- (a) 男性と女性の別々の試験:試験間の共変量としての性別. これは A1 と同等であり解析方法は 3.7.1 節で説明する.
- (b) 男性と女性の両方を含む試験で、男性と女性の割合を報告しているが、性別ごとアウトカムの推定値(不確実性を含む)が報告されていない.この割合は、時には、試験レベルの連続した共変量(女性の割合)と見なされる.このタイプのメタ回帰の解析方法は、3.7.3節で説明する.
- (c) 男性と女性の両方を含む試験だが、性別ごとの割合又は性別ごとのアウトカムが報告されていない。 さらなる前提条件がない限り、メタ回帰は実行できない.

A3. 患者レベル共変量

- (a) 関心のあるアウトカム及び共変量の IPD が利用可能な試験. この場合, 試験内での共変量の効果の検討に利用でき, その後メタ回帰でさらに検討できる.
- (b) 例えば、男性と女性の両方を含む試験だが、グループごとに個別に、治療効果とともに精度の指標 (バラツキ) が報告されている場合. これは試験内の効果であり、メタ回帰の目的では、性別に 関しIPDを有することと同等である.この同等性はアウトカムが2値又は連続値であれば正しく、カテゴリの共変量にのみ適用される.

次に,連続値の共変量も,同様の区別ができる.

B1. 試験特性に関連する試験レベル共変量

例えば、薬物の投与量、解析方法は3.7.3節と3.7.4節で説明する.

B2. 患者特性に関連する試験レベル共変量

例えば,試験中の患者の平均年齢. これは B1 と同等であり,解析方法は 3.7.3 節と 3.7.4 節で説明する.

B3. 患者レベル共変量

アウトカムが 2 値の場合, 平均年齢と不確実性の指標※が, 事象あり及び事象なしで別々に報告されていれば,メタ回帰の目的のためには,各患者の正確な年齢を記録した IPD を有することと同等である. 共変量の平均値が別々に報告されない場合,メタ回帰を行うために IPD が必要となってしまう. 連続値の共変量で,連続値のアウトカムの場合は,常にメタ回帰では IPD が必要とされる. これは 3.7.5 節で説明する.

※補足:他の節では精度の指標(バラツキ)として参照しているが、本節では試験間の差異などを指し示す用語として不確実性と参照されていると考えられ、費用対効果評価では「不確実性」として参照されることが多いため、そのまま使用した.

治療と共変量との交互作用を調べる場合,ある共変量の値での治療効果を,共変量の他の値での治療効果とを比較することになる. 試験内での群間比較(試験内比較)と複数の試験を用いた群間比較(試験間比較)には2つの重要な相違がある. 1つ目の相違点は、性別のようなカテゴリ共変量では、試験内比較と試験間比較との違いは、対応のあるt検定と対応のないt検定の違いと非常によく似ている. 試験間比較では、所与の共変量の効果(すなわち交互作用)は、試験間のバラツキにより生成される「ランダムノイズ」と区別可能でなければならないため、検出するのがより困難になる. しかし、試験内比較には、交互作用はサンプリングエラー(標本誤差)と区別可能であることのみが必要となる. 各試験内の精度は比較的高いかもしれないが、試験間比較では観測数(試験数)は非常に少なくなることが多いため、治療効果と共変量との間に統計的に高度に有意な関係性を検出する可能性が非常に高いことである(偶然による偽陽性の交互作用を検出しやすくなる)[10].

3.2.2 生態学的誤謬(ごびゅう)(Ecologic Fallacy)

2 つ目の相違点は、試験間比較がエコロジカルバイアス又は生態学的誤謬に対して脆弱なことである [11]. 例えば、試験間レベルの共変量に対する治療効果の線形回帰係数が、試験内データから得られた 回帰係数と全く異なり得る現象である. つまり、試験レベルの共変量と治療効果がメタ回帰等では成立しているように見えても、患者レベルの共変量と治療効果との間には実際には成立していない現象である. もちろん、同一のエビデンス統合で、試験内情報 A3 (b) 及び試験間情報 A2 (a) の両方を有することは、十分に可能である. 連続値の共変量では、すべてのデータが IPD (B3) である場合、共変量の平均値に基づく試験間の回帰係数と、共変量の平均値の周りの個々の変動に基づく試験内の回帰係数の両方を推定するモデルを適用することは可能である. IPD の解析方法は 3.7.5 節で議論する.

3.2.3 連続値の共変量を伴う個別患者データ (IPD) を用いた群間比較でのより大きな検出力

連続値の共変量と IPD を用いると、試験内比較ではエコロジカルバイアスを回避できるだけでなく、 共変量の真の効果に対する統計学的検出力がはるかに大きくなる.これは患者の共変量の値の変動 (バラツキ)が、試験の平均値間のバラツキよりも何倍も大きく、メタ回帰で推定された回帰係数の精度は、 共変量の分散に直接的に影響を受けるためである.

3.2.4 折りたたまれたカテゴリデータの利用(Use of collapsed category data)

A2 (b) の状況は、あたかも共変量が報告されていないかのように、治療効果が統合されたデータからのみ計算されている状況で、「折りたたまれた」カテゴリデータとして参照される[12][13]. これらの場合、試験内での群間比較はできるが、データとしては不完全である. 非常に一般的に見られるデータ構造は、幾つかの試験は男性で実施され、幾つかの試験は女性で実施され、第3のカテゴリは男性と女性の割合が報告されている試験で、これらが混在している. 試験間比較での共変量として男性の割合を用い、これらの試験を単一の解析に統合させることが可能である. 共変量は、男性の試験では1、女性の試験では0とする. そのようなデータは、サブグループの解析(3.7.1 節)の方法論を用いて解析できる. しかし、そのモデルは、直線的なモデル、言い換えれば恒等式(identity)のリンク関数を持つモデル(TSD2[8]参照)に対してのみ厳密に正しいことに注意する. 一方で、logit、log、又は他の一般的に使用されるモデルでは妥当ではない[11].

非線形モデルを用い、これらすべてのデータ形式(割合、カウント、連続データなど)に対し、折りたたまれたカテゴリを取扱う方法がある。これは TSD の範囲を超えているが、公表論文を参照でき、その考え方はこの問題を解決するために適用できる[12][13][14]。これらの方法は、共変量の情報が完全に「欠測」している状況である A2 (c) タイプの試験のデータを取込むように更に拡張できる。A2 (c) タイプのデータに対する治療効果の推定に対しては説明されていないが、同様の応用事例[12][13]の考え方やプログラミングコードが適用可能である。

3.2.5 要約によるバイアス (Aggregation bias)

最後に、共変量と治療効果との交互作用がないが、共変量がベースラインリスクを変化させる場合(重症な患者集団では進行が早くなる等)、共変量に対するデータを統合する影響は、推定された治療効果を無効(方向)に偏らせることと理解する必要がある。これは要約することによるバイアス(要約バイアス)[11]として知られるエコロジカルバイアスの一形態であり、直線的なモデルに厳密には影響しない。そのような共変量のデータを統合することではバイアスが生じない。

一方で、通常は生存時間解析では、共変量がベースラインリスク及び治療効果の両方に強い影響を及ぼす場合にのみ重要であり、年齢などの共変量の癌のベースラインリスクへの影響が特に顕著であり、且つルーチン的に使用される対数線形モデル(補足:ここでの線形とは、log リンク変換された反応変数に対してモデル化する場合、右辺をパラメータの線形結合で説明できるという意味である)は非常に非線形であるため、特に危険性がある。したがって、非線形のモデルにおいては、ベースラインリスクに影響する共変量が存在する場合、例えそれらが治療効果を変化させないとしても、統合データから推定

された治療効果に基づいてではなく、グループ(カテゴリ)共変量での層別解析、及び連続値の共変量の回帰分析からの治療効果を統合した推定値に基づかなければならない.

3.3 バイアス(偏り)の調整方法の概要

バイアス調整の目的は、対象母集団での所望の効果に関して偏った治療効果の推定値を、効果的に不偏推定値に変換することである。3.8 節では、4 つの方法について論じるが、そのうちの 2 つはメタ回帰の種類である。すべての場合で、外部データ又は使用されている事前の意見の不確実性を考慮する必要がある。4 つの方法には、外的妥当性を脅かすバイアス(以下、外的バイアスと呼ぶ)の共変量による調整方法(3.8.1 節)、内的妥当性を脅かすバイアス(以下、内的バイアスと呼ぶ)のための外部データに基づく、バイアスリスクのあるエビデンスの調整及び重みづけを小さくする方法(3.8.2 節)、NMAの内的バイアスのリスクのマーカーに関連するバイアスの推定(3.8.3 節)、及び専門家の意見や他のエビデンスに基づいて内部並びに外部のバイアスを調整する方法(3.8.4 節)が含まれる。

3.4 メタ回帰の形式としてのネットワークメタアナリシス

確かに、NMA はメタ回帰の一形態として理解できるが、無作為化の比較に基づいていることを強調すべきである[15]. 実際、NMA で一致性を仮定した治療効果の推定値(TSD2[8]参照)は、2 群比較のメタアナリシスの場合と同様に、個々の試験からの推定値に対する重み付き推定値である[16]. また、NMA 又は間接比較は、観察研究を含めることによるバイアスに脆弱とするのは少し誤解を生みやすい[3]. 2 群比較のメタアナリシスと同様に、NMA 又は間接比較は、文献等から収集できない効果修飾因子の問題を抱えていることが問題である(補足:未測定の交絡など). NMA 又は間接比較は、構成する試験が、その対象集団に対して不偏である限り、治療効果の不偏推定値を与える. したがって、NMA 又は間接比較 比較を構成する試験は無作為化の比較に基づいているため、観察研究よりも優れている.

3.5 異質性の測定

試験間の異質性を検討する多くの標準的方法が提案されており、読者は標準的な教科書[3][9][17]を参照されたい. 文献では、固定効果(Fixed Effects; FE)モデルの均一性に関する帰無仮説の検定は、例えば、コクランの Q 統計量を用いて、しばしば変量効果(Random Effects; RE)モデルの選択を正当化するために使用される. I²統計量は、尺度フリーであるという利点があるが、含まれている研究の数及びそのサイズ(各試験のサンプルサイズ)に依存しており、典型的なメタアナリシス[18]では解釈が難しい. TSD2[8]でのアプローチ、つまりベイズの枠組みでは、固定効果と変量効果のモデルの残差逸脱度と DIC(Deviance Information Criterion)統計量を比較することであった[19]. ベイズのアプローチの利点は、それが試験間の分散の事後分布を与えることである一又は恐らく解釈がより簡単であること一治療効果の試験間での標準偏差(治療効果の真の分布における標準偏差)は、データと互換性のある値の範囲に関する何らかの洞察を研究者に与える(補足:将来のデータも含めて、どの程度ばらつくかの知見が得られる). 古典的アプローチ[20]を用い、試験間分散の不確実性の尺度(測定値)を得ることも可能であるが、これは頻繁には行われない. しかし、TSD2[8](6.2 節)に与えられている重要な警告を繰り返す

必要がある. 試験間の標準偏差の事後分布が、事前分布に非常に敏感(センシティブ)である可能性があり、特に、漠然事前分布(vague priors)を用いるという「デフォルト」は、非現実的に高いレベルの異質性を許す事後分布をもたらす. 試験数が少ないとき、又は試験の大部分のサンプルサイズが小さいときに必ず発生する. 解決策は、専門家の意見又はメタ疫学研究(補足:試験デザインの特徴が治療効果にどのように影響を及ぼすかを検討する研究)からの情報に基づいて、無情報ではない事前分布(informative prior)を利用することである. (無情報ではない事前分布を得るための)最も簡単なアプローチは、同じ条件で、同じアウトカム指標を採用している他の治療法での大規模なメタアナリシスを特定することかもしれないし、このメタアナリシスから得られる試験間の異質性の事後分布を用いて、現在の解析に利用することかもしれない[21].

3.5.1 意思決定での異質性の関わり合い

これまで比較的注意を払わなかった重要な問題は、費用対効果評価の意思決定において、高いレベルの異質性にどの様に対応するかである。研究者が治療効果の大きさを試験間のバラツキの程度と比較することが重要である。図 3-1 は、変量効果モデルが適合した状況を示している。平均治療効果の事後平均は 0.70 であり、事後分布の標準偏差(sd)= 0.2 であり、平均治療効果(95%信用区間(0.31,1.09))は明らかに 0 を含んでいない。しかし、試験間の標準偏差の事後平均は σ =0.68 であり、平均効果と同様の大きさである。次に、試験数(サイズ)が無限大の将来の試験のアウトカムを予測する合理的な信頼区間は何か?という疑問を考察する。古典的な統計学での近似解は、平均値の分散を試験間の分散に加えて、sd²+ σ ²=0.50 から予測分布の標準偏差 0.71 を与えることである。95%予測区間は、有害な影響(補足:効果が対照薬に負けている方向)を含み、効果がない部分に簡単に及ぶ(-0.69、2.09)ことに注意する。これらの分布をベイズの方法で解釈すると、平均効果がゼロ未満となる確率は 0.0002 であるが、一方で新しい試験が負の治療効果を示す確率ははるかに高い(0.162)(図 3-1) である。

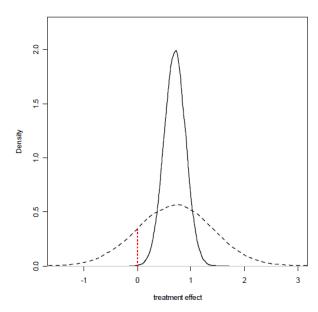


図 3-1 事後平均=0.7, 標準偏差=0.2, 試験間の標準偏差(異質性)の事後平均=0.68(補足:事後平均 0.7 と同程度)の治療効果に対する事後確率密度(実線)及び予測確率密度(破線).

[出典: NICE DSU TSD3 (http://nicedsu.org.uk/technical-support-documents/technical-support-documents/)]

この問題は幾つかの文献で考察されており[6][22][23][24], 異質性の存在下では, 平均治療効果の(事後)分布よりむしろ予測分布が, 将来「ロールアウト」(拡大的に売り出していく)する特定の介入の治療効果の比較に対する不確実性をよりよく表すことが示唆されている. ベイズのマルコフ連鎖モンテカルロ法 (Markov Chain Monte Carlo; MCMC)の設定では, 治療効果の分布から更なる標本を引き出すことによって, 予測分布を容易に得ることができる.

$$\delta_{New} \sim N(d, \sigma^2)$$

ここで、d は推定された (共通) 平均治療効果であり、 σ^2 は推定された試験間の異質性の分散 (TSD2[8] 参照) である.

線形尺度での予測分布の平均は、平均効果の分布の平均と同じになる.しかし、決定の不確かさに対するこの推奨の影響は、説明できない異質性が高レベルである場合には、かなり重大である可能性があり、したがって異質性の程度は誇張されないことが重要である.これは直ちに、異質性の原因が何であるか?という疑問を直ちに提起する[25][26].それは引き続く節でより詳細に取り上げ、想定される原因に関して、試験結果を調整することで異質性を減じる方法を論じる.現在の目的のために、患者集団全体にわたる治療効果の大きさの真のバラツキと、試験が行われた方法により引き起こされたバイアスによる見かけ上のランダム変動とを区別する. Higginsら[6]は、予測分布の分散項は、対象集団間の真のバラツキのみで構成されるべきことを明確にしている。しかし現時点では、様々な変動原因を区別できる明確な方法論や情報源は存在しない。最近の非常に多数のメタアナリシスのメタ疫学研究ではこのことを明らかにし始めているが、現時点で確信を持って言えることは、観察された異質性は、治療効果の大きさの真のバラツキが過大評価されていそうという点である。

この議論では、含まれる全ての試験で治療効果が交換可能性があると仮定されている。しかしながら、 意志決定に関わる対象集団は、幾つかの試験の対象集団に類似しているかもしれない(その他の試験と はあまり類似していないかもしれない).この場合、外的妥当性に対する調整を考慮すべきである(3.8 節参照).

3.6 外れ値の検出

異質性 (heterogeneity) の疑問に密接に関係するのは、外れ値の検出の方法である。ここでの焦点は、試験結果の全体の変動ではなく、他の試験から異なる結果を示す1つ又は2つの試験である。中心から離れた1つの試験が、異質性の測定値に大きな影響を与える可能性があるため、この2つの問題は密接に関係している。逆に、異質性の水準が高いと真の外れ値を検出するのが難しくなる。

3.6.12群比較のメタアナリシスでのクロスバリデーションによる予測の妥当性評価

急性心筋梗塞の患者を対象にしたプラセボ対照の静脈内マグネシウム投与の 16 試験[27]について、表 3-1 のデータから計算される未調整の対数オッズ比、及び変量効果モデルからの「縮小」推定値(すなわち、交換可能であると仮定したもとでの試験ごとの治療効果)のフォレストプロットを図 3-2 に示す.

表 3-1 急性心筋梗塞の患者でのプラセボに対する静脈内マグネシウム投与の 16 試験[27]での患者数と死亡例数

			Plac	ebo	Magnesium	
Trial ID	Trial Name	Year	Deaths	Total	Deaths	Total
1	Morton	1984	2	36	1	40
2	Rasmussen	1986	23	135	9	135
3	Smith	1986	7	200	2	200
4	Abraham	1987	1	46	1	48
5	Feldstedt	1988	8	148	10	150
6	Shechter	1989	9	56	1	59
7	Ceremuzynski	1989	3	23	1	25
8	Bertschat	1989	1	21	0	22
9	Singh	1990	11	75	6	76
10	Pereira	1990	7	27	1	27
11	Shechter1	1991	12	80	2	89
12	Golf	1991	13	33	5	23
13	Thorgersen	1991	8	122	4	130
14	LIMIT-2	1992	118	1157	90	1159
15	Shechter2	1995	17	108	4	107
16	ISIS-4	1995	2103	29039	2216	29011

[出典: NICE DSU TSD3 (http://nicedsu.org.uk/technical-support-documents/technical-support-documents/)]

本データに対しては、残差逸脱度 (residual deviance) の事後平均77.5,及びDIC=94.5である固定効果モデルに対し、残差逸脱度の事後平均29.6 (32 (データポイント)と同程度)及びDIC=54.2 (<94.5)の変量効果モデルが選択された(より詳細はTSD2[8]参照).対数オッズ比の尺度での標準偏差の事後分布の中央値は0.68 (95% 信用区間 (0.35, 1.30))で、治療効果(平均値)の大きさ -0.89 (95% 信用区間 (-1.49, -0.41))と同程度である。これは顕著な異質性があることを示している。

図 3-2 では、ISIS-4 大規模試験の対数オッズ比が 0.055 (95% 信用区間 (-0.007, 0.117))と推定され、他の試験とは多少異なることが示されている. 特に、本試験の未調整 95%信頼区間 (CI) も「縮小」95%信用区間(補足:全体の結果も考慮することで、各試験の未調整の信頼区間よりも縮小している)も、平均治療効果の 95%信用区間と重なっていない (図 3-2). 研究者は本試験がある種の意味において「外れ値」ではないかと疑うかもしれない. メタアナリシスにおいて、1 試験ごとの異質性を調べる適切な方法は、「1 つずつ除く(leave one out)」アプローチに基づくクロスバリデーション[28][29]である. その手順は、まず本試験を統合する対象から除き、次に残りの試験に基づいた解析で期待される効果の予測分布と、観測された治療効果を比較する. クロスバリデーションによる予測の妥当性評価における最初のステップは、試験 16 の ISIS-4 を除いた表 3-1 のデータに変量効果(メタアナリシス)モデルをあてはめることである.

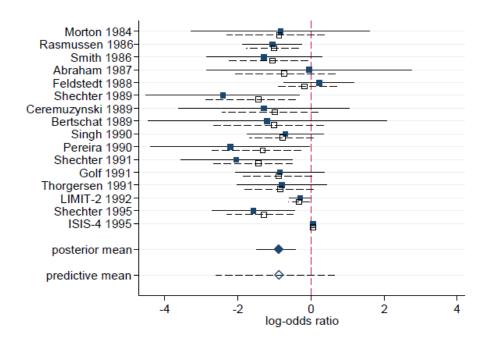


図 3-2 マグネシウムの例:未調整の対数オッズ比と 95%信頼区間(■と実線);試験ごとの対数オッズ比(95%信用区間)の「縮小」推定値(□と破線);全ての試験を含む変量効果モデルから得られた,統合した治療効果の事後平均と 95%信用区間(◆と実線)と予測分布の 95%信用区間(◇と破線)

[出典: NICE DSU TSD3 (http://nicedsu.org.uk/technical-support-documents/)]

TSD2[8]の表記に従い, r_{ik} はイベント数(死亡数)を表し,試験 i,群 k での各群の患者数 n_{ik} からの二項尤度 $r_{ik}\sim Binomial(p_{ik},n_{ik})$ に従うと仮定する.ここに p_{ik} は,i=1,...,15 (試験 16 の ISIS-4 を除く);k=1,2 について試験 i の群 k におけるイベント確率(イベント割合)を示す.変量効果モデルは,

$$logit(p_{ik}) = \mu_i + \delta_{i,1k} I_{\{k \neq 1\}}$$

ここで,

$$I_{\{u\}} = \begin{cases} 1 & u$$
が真 0 その他 (1)

である. 試験ごとの対数オッズ比は共通の分布: $\delta_{i,12} \sim N(d,\sigma^2)$ に従う. 次のステップは、予測分布から将来の試験で予測される治療効果 δ_{new} を抽出することである.

$$\delta_{new} \sim N(d, \sigma^2)$$

ここで、d と σ は事後分布から得られる.次に ISIS-4 と同じベースラインリスクと,同じ規模の試験 から繰り返し抽出する必要があり,予測治療効果 δ_{new} にそれらを適用する.本試験でのベースライン効果は,プラセボの死亡割合 p_{base} の logit 変換したもので,ISIS-4 のプラセボ群の死亡割合 (2103/29039=0.072) から推定できる.しかしながら,これは死亡割合の不確実性を考慮していない.その代り,新たな試験(ISIS-4)での死亡割合が,以下の Bata 分布に従うと仮定することができる.

$p_{base} \sim \text{Beta}(a, b)$

ここに、 $a=r_{16,1}=2103$ は、試験16(ISIS-4)のプラセボ群のイベント数であり、 $b=n_{16,1}-r_{16,1}=26936$ は、試験16のプラセボ群でのイベントを発現しなかった患者数である(Bate分布の期待値:a/(a+b)).残りの15試験が所与の下、将来の試験(ISIS-4)でのマグネシウム群の死亡の予測確率(予測される死亡割合) p_{new} は、

$$logit(p_{new}) = logit(p_{base}) + \delta_{new}$$

で与えられる. 試験16 (ISIS-4) と同じ規模の将来の試験でのマグネシウム群の予測イベント数 r_{new} は、死亡確率 (死亡割合) p_{new} をパラメータとする以下の二項分布

$$r_{new}$$
~Binomial $(p_{new}, n_{16.2})$

から抽出される.ベイズ流のP値,すなわち試験16で観測されたイベント数と同じくらい極端な値を得る確率 $\Pr(r_{new} > r_{16,2})$,を得るために,試験16(ISIS-4)での観測イベント数と比較する.ベイズ流MCMCの枠組みでは,これは各繰返し(iteration)で, $r_{new} > r_{16,2}$ であれば1をとり,それ以外は0をとる変数を設定し,その繰返しの平均を求めることで(補足:1をとった数を総繰返し数で割る),ベイズ流でのP値が算出できる.

オリジナルの変量効果モデルのWinBUGSコードは, TSD2[8] (Program 1(a) or (c))に与えられている. 以下に予測的なクロスバリデーションのWinBUGSコードを示す. プログラムコードは, 強調された部分を除いて, TSD2[8] (プログラム1 (a)) に示された2群のメタアナリシスのコードと同じである.

プログラム1: Binomial likelihood, logit link, 予測的なクロスバリデーション, 2つの治療法(マグネシウムの事例).

```
# Binomial likelihood, logit link, 2群のメタアナリシス (2治療)
# 変量効果モデル, 予測的なクロスバリデーション
                                                 # *** PROGRAM STARTS
model{
                                          # 試験のループ処理
for(i in 1:ns){
                                          # 対照群の治療効果は0
delta[i,1] <- 0
mu[i] \sim dnorm(0,.0001)
                                                 # 試験ベースラインの漠然事前分布
for (k in 1:2) {
                                                 # 治療群のループ処理
   r[i,k] \sim dbin(p[i,k],n[i,k])
                                                 # 二項尤度
                                                 # 線形予測に対するモデル
   logit(p[i,k]) <- mu[i] + delta[i,k]
                                                 # 分子の期待値
   rhat[i,k] \leftarrow p[i,k] * n[i,k]
   dev[i,k] \leftarrow 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
                                                 # 逸脱度
   + (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
                                                 # 試験ごとの残差挽脱度の合計
resdev[i] <- sum(dev[i,])
delta[i,2] ~ dnorm(d[2],tau)
                                                 # 試験特有の対数オッズ比(LOR)の分布
totresdev <- sum(resdev[])
                                                 # 残差逸脱度の総合計
d[1] < 0
                                                 # 参照治療1の治療効果は0
d[2] \sim dnorm(0,.0001)
                                          # 治療効果の漠然事前分布
```

 $sd \sim dunif(0,5)$ # 試験間標準偏差の漠然事前分布 tau <- pow(sd,-2)# 試験間精度=(1/試験間分散) delta.new ~ dnorm(d[2],tau) # 将来の試験の予測分布 p.base ~ dbeta(a,b) # ベースライン (対照群) の効果の生成 a <- r[ns+1,1] # 対照群のイベント数 b <- n[ns+1,1]-r[ns+1,1] # 対照群の非イベント数 logit(p.new) <- logit(p.base) + delta.new</pre> # 治療群の予測イベント率 r.new ~ dbin(p.new, n[ns+1,2]) # 治療群の予測イベント数 # ベイズ流のp値: 得られた値より極端な値を得る確率 # モデル及び残りのデータを与えられたもとでの観察されたイベント数 (r[ns+1,2])との比較 p.cross <- step(r.new - r[ns+1,2]) - 0.5*equals(r.new,r[ns+1,2]) # より大きな極端な値の割合 # *** PROGRAM ENDS

クロスバリデーションの P 値は、p.cross を検討することで得られる。関心のある他のパラメータの事後分布からの要約を得るには、d、delta.new、及び sd を検討する必要がある。モデルの適合度及びモデルの比較を評価するために必要なパラメータの事後平均値を得るために、dev、totresdev、及び (WinBUGS DIC ツールから得られる) DIC を検討する必要がある。更に、図 3-2 や図 3-3 のような「縮小」推定値のプロットを作成するには、delta をモニターする必要がある。

データ構造は TSD2[8]で示したデータ構造と同一だが、データの最終行の試験(本事例では ISIS-4 試験)に対して、クロスバリデーションの P 値を計算する。簡単に説明すると、ns は試験数で、データ本体部分では、r[,1]及び n[,1]は最初の治療群の分子及び分母であり、r[,2]及び n[,2]は 2 番目の治療群の分子及び分母であり、そして除外される試験は最後の行に与えられている。

データと連鎖ごとの初期値

```
# Data (Magnesium Example)
list(ns=15)
                        # ID
r[,1] n[,1]
           r[,2]
                 n[,2]
    36
           1
                 40
                        # 1
23
    135
          9
                 135
                        #2
         2
    200
                 200
                       #3
(中略)
                 107
                        # 15
17 108
2103
          29039 2216 29011 #16
FND
# Initial values
# Initial values for delta and other variables can be generated by WinBUGS.
list(d=c( NA, 0), sd=1, mu=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0), p.base=0.5)
#chain 3
list(d=c( NA, 2), sd=2, mu=c(-3,5,-1,-3,7, -3,-4,-3,-3,0, -3,-3,0,3,5), p.base=.8)
```

その結果、P値は0.056となり、残りの試験データに対するモデルが所与の下(20,000回のBurn-inの後、収束し、3つの独立連鎖からの50,000回のサンプルに基づく結果である)、ISIS-4と同じくらい極端な結果は起こりそうもないことが示唆されたものの、可能性はまだある(補足:恐らく5%を上回っているた

め). しかしながら、これらの結果を検討する場合、実施できる検定の有効数を考慮しなければならない. ISIS-4に対するクロスバリデーションを実行する際、16試験で最も極端な試験を取り上げた. すなわち、検定はn=16回実施され、そのうちISIS-4に対する検定が最も極端な結果(すなわち、最も小さなP値)を与える. その場合、観測されたP値の有意性を正確に解釈するためには、そのP値をその期待値、すなわちn番目の一様順序統計量の値である1/(n+1)=0.059と比較する必要がある※. その結果、観測されたP値は、ISIS-4試験以外の試験データに当てはめた変量効果モデルと、ISIS-4試験が必ずしも適合していないわけではないことを示唆している. このことは、ISIS-4試験で観測されたオッズ比と信頼区間、ISIS-4試験を除いた場合の変量効果メタアナリシスでの「縮小」推定値、治療効果の事後平均及び予測平均が示されている図 3-3でも確認できる. それはISIS-4試験から観測された対数オッズ比は、治療効果の事後平均の信用区間からは外れているが、予測的なクロスバリデーションの基本となる治療効果の予測平均の信用区間の範囲内に依然として存在していることが見てとれる. これは統計的な結果に過ぎない. すなわち、ISIS-4試験が逸脱した結果であるか、あるいはそれ以外の試験が逸脱しているのかを、推論することは不可能である. 本メタアナリシスは繰り返し議論され[30][31]、現在の意見ではISIS-4は実際「正しい」結果である[32]とされている.

※補足:ここでのクロスバリデーションの P値とは,ある 1 試験を除いて,残りの試験に基づくモデルにより推定を行い,当該 1 試験のマグネシウム群の死亡数と,当該 1 試験と同じ規模の試験に対して,モデルから予測された死亡数の方が大きくなる確率として計算される.その場合,帰無仮説(試験間に違いがない)のもとでは P値は一様分布に従うことより,一様分布 U(0,1)から無作為に n 標本(試験数)を抽出することになる.一様分布から抽出された n 個の値を,小さい順に並べて k 番目の値(確率変数)の確率分布は,ベータ分布 B(k,n-k+1) に従うことが知られており,その期待値は k/(n+1) (、ベータ分布 B(a,b)の期待値は a/(a+b))となることから,最小値 (小さい方から 1 番目) の期待値は 1/(n+1) となる.

事例では 16 試験を 1 つずつ除いて検定した場合、対応する 16 個の P 値が得られ、帰無仮説(試験間差はない)のもとでは P 値は一様分布し、その場合一番小さな P 値の取りうる値は、最小値の期待値であることから 1/(16+1)=0.059 となる。したがって、ISIS-4 試験を除いた場合の P 値は 0.056 であることから、最小値の期待値よりは小さいものの大きく下回るものではなかったため、適合しないとまでは言えないことになる。

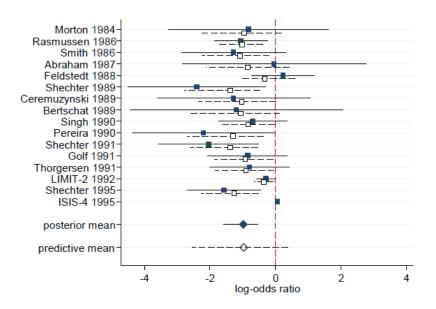


図 3-3 マグネシウムの例:未調整の対数オッズ比と 95%信頼区間(■,実線). 試験ごとの対数オッズ比の事後平均と 95%信用区間,すなわち「縮小」推定値(□,破線). ISIS-4 試験を除いた変量効果モデルに基づく,統合した治療効果の事後平均と 95%信用区間(◆,実線)及び予測分布の 95%信用区間(◇,破線)

[出典: NICE DSU TSD3 (http://nicedsu.org.uk/technical-support-documents/)]

本マグネシウムのデータセットには、意思決定における変量効果モデルに関する幾つかの重要なメッセージがある. 1 つ目は、ISIS-4 試験を含む変量効果モデル、ISIS-4 を含まない変量効果モデルは、同程度によく当てはまったことである(2 つのモデルの残差逸脱度の事後平均は、データポイント 32 と 30 に対して、それぞれ 29.7 と 27.9 であった). これは、変量効果モデルは、一般的に治療効果の任意の確率分布に適用でき、その広がりによって大きくは影響を受けないためである。2 つ目として、平均効果に基づいて推定することの弱点を示している。全体の試験の組合せの中で考えると、ISIS-4 試験自体を含めたり、除いたりしても、特に注目すべきものではない。しかしながら、ISIS-4 試験は(統合された)平均効果からは著しく異なっている。したがって、治療効果の平均効果に基づく意思決定は、様々なエビデンスソースにおいて、説明できない不一致がある場合のモデルに基づく意思決定となる。(一方で)治療効果の予測分布に基づくモデルは、全てのデータと矛盾しない。

3.6.2 間接比較及び NMA による予測の妥当性評価のためのクロスバリデーション

クロスバリデーションは、複数の治療法と複数群で実施された試験で実施されるより広いエビデンスのネットワークに対して、その方法/考え方を修正することなく適用できる。しかしながら、複数の治療法がある場合、予測分布は多変量正規であることを念頭におく必要がある。そのため、s個の治療法のネットワークに対し、治療1 [基礎パラメータ(the basic parameter)、TSD2[8]を参照]に対するs-1個の治療効果の予測分布は、以下で与えられる。

$$\delta^{new} = \begin{pmatrix} \delta_{12}^{new} \\ \vdots \\ \delta_{1s}^{new} \end{pmatrix} \sim N_{s-1} \begin{pmatrix} d_{12} \\ \vdots \\ d_{1s} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2/2 & \cdots & \sigma^2/2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2/2 & \sigma^2/2 & \cdots & \sigma^2 \end{pmatrix} \end{pmatrix}$$
(2)

ここに、dと σ は(データが与えられた下で)事後分布から抽出される.この分布は、一連の条件付き 1変量正規分布として書き直すことができる[33].

$$\delta_{1k}^{new} | \begin{pmatrix} \delta_{12}^{new} \\ \vdots \\ \delta_{1(k-1)}^{new} \end{pmatrix} \sim N \left(d_{1k} + \frac{1}{(k-1)} \sum_{j=1}^{k-1} (\delta_{1j}^{new} - d_{1j}), \frac{k}{2(k-1)} \sigma^2 \right)$$
(3)

式(2)の多変量分布又は式(3)の条件付分布のどちらかは、治療法1(参照治療)に対する各治療法の予測された変量効果を推定するのに用いられなければならない。治療法の数に依存しないネットワークを構成できる、より一般的なコードを可能にするため、以下に示すWINBUGSコードはTSD2[8]にしたがい、式(3)の形式を利用する.

予測的な治療効果間の相関が、治療間の全ての対比に正確に反映されることを保証するため、他の治療比較(参照治療群以外との比較)に対する予測分布は一致性の式(TSD2[8])から得られている.

$$\delta_{XY}^{new} = \delta_{1Y}^{new} - \delta_{1X}^{new}$$

WinBUGS プログラムコードを以下に示す.化学療法中の有害事象である熱性好中球減少症(FN)の発現率を低下させる3つの治療法に関するエビデンス統合が,費用対効果評価のために実施された(Madan, J. et. al., Value in Health 2011).「無治療」は参照群として1とコード化され,関心のある3つの治療法であるフィルグラスチム(Filgrastim),ペグフィルグラスチム(Pegfilgrastim)及びレノグラスチム(Lenograstim)は、 $2\sim4$ にコード化される.

表 3-2は試験の各群の全患者 n_{ik} のうちFNを発現した患者数 r_{ik} , 及び治療法 t_{ik} を示している(i = 1, …, 25; k = 1,2). 図 3-4にネットワークダイアグラムを示す.

表 3-2 化学療法を受けた患者の合計 n_{ik} のうち、(無治療も含む)4 治療 t_{ik} に対する 25 試験の群 1 及び群 2 での有害事象の発現数 r_{ik} .

	治療		有害事绩	象の例数	化学療法	を受けた 例数
	群 1	群 2	群 1	群 2	群1	群 2
試験 ID	t_{i1}	t_{i2}	r_{i1}	r_{i2}	n_{i1}	n_{i2}
1	2	3	15	10	75	77
2	2	3	27	14	147	149
3	2	3	2	5	25	46
4	2	3	6	6	31	29
5	2	3	1	0	13	14
6	1	2	26	34	72	276
7	1	2	17	9	39	41
8	1	2	15	4	72	77
9	1	2	86	72	192	197
10	1	2	52	34	104	101
11	1	2	62	40	125	125
12	1	2	27	16	85	90
13	1	2	80	38	104	95
14	1	2	34	17	64	65
15	1	2	38	25	130	129
16	1	4	18	5	28	23
17	1	4	42	36	59	61
18	1	4	15	5	26	22
19	1	4	62	52	80	82
20	1	4	14	5	43	43
21	1	3	27	11	73	73
22	1	3	34	14	343	343
23	1	3	5	4	29	30
24	1	3	10	3	118	123
25	1	3	78	6	465	463

[出典: NICE DSU TSD3 (http://nicedsu.org.uk/technical-support-documents/technical-support-documents/)]

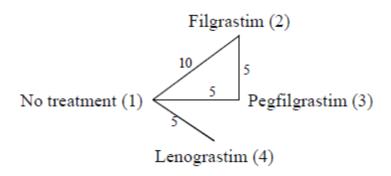


図 3-4 化学療法における有害事象:治療ネットワーク.2つの治療法を結ぶ線は、これらの治療法間の比較が行われたことを示す。線上の数字は、2つの治療法を比較した RCT の数を示している。() 内は治療番号を示す。

[出典: NICE DSU TSD3 (http://nicedsu.org.uk/technical-support-documents/)]

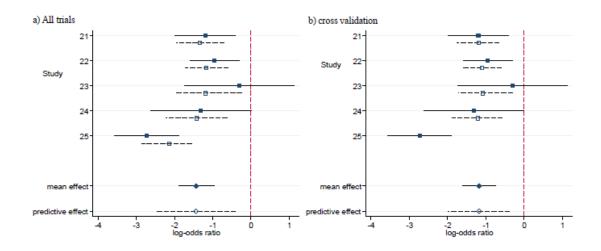


図 3-5 化学療法における有害事象: 粗対数オッズ比及び95%信頼区間(黒四角,実線); 試験ごとの対数オッズ比の事後平均値及び95%信用区間, すなわち「縮小」推定値(白四角,破線); a)全ての試験を含めた変量効果モデル,及びb)25番目の試験を除いた変量効果モデル(クロスバリデーションモデル)を用いて,統合した対数オッズ比の事後平均値及び95%信用区間(黒菱形,実線)及びその予測分布(白菱形,破線)

[出典: NICE DSU TSD3 (http://nicedsu.org.uk/technical-support-documents/)]

図 3-5 (a) に、治療法1と治療法3を比較した試験に対するデータから計算した粗対数オッズ比と、「縮小」推定値(すなわち、交換可能を仮定した試験ごとの治療効果)のフォレストプロット、及び表 3-2 のすべての試験を含む変量効果モデルから治療法3と比較した治療法1の事後的効果及び予測的効果を示す。変量効果モデルでのNMAは、データによく適合するが、試験25の推定された対数オッズ比は他の試験とは少し異なっており、ネットワークにおける高度な異質性に寄与している可能性がある(σ =0.42 の事後中央値及び95%信用区間(0.20,0.73))。試験25が「外れ値」であるかを検討するため、前節で説明したように「1つ除外する」アプローチに基づくクロスバリデーションを利用した。P値は0.004であり、試験25以外のデータに対するモデルを与えたもとでは、試験25ほど極端な結果になる試験は、ありそうもないことを示している(60,000回のBurn-in後に収束し、結果は3つの独立した連鎖(Chains)からの100,000サンプルに基づく)。

標準的な変量効果モデルに適合する WinBUGS プログラムコードは, TSD2[8] (プログラム 1 (c)) に 与えられている. NMA における予測的なクロスバリデーションのための WinBUGS プログラムコードは, プログラム 2 で与えられている. 以下のプログラムコードは一般的であり, 2 群のメタアナリシス及び (複数群試験の有無にかかわらず) NMA における予測的なクロスバリデーションに利用できる.

これは 25 試験のうち最も極端な試験を選択し、すなわち(直接比較又は間接比較により) 25 回の検定を行えるという意味である。実行できる検定の有効数 (n) を考慮し、観測された P 値をその期待値 (1/(n+1)=0.038、すなわち一様順序統計量の値)と比較する必要がある。観測された P 値はこれよりかなり小さく、試験 25 が「外れ値」である可能性を示している。

図 3-5 (b) は、治療法1と治療法3を比較した試験での「縮小」推定値の平均値及び予測的な治療効

果,及び試験 25 を除いた変量効果モデルにおける事後的な治療効果及び予測的な治療効果を示している (ただし,試験 25 で観察された対数オッズ比と信頼区間を含む). 試験 25 で観察された対数オッズ比の 95%信頼区間は (-3.57,-1.89) で,事後平均の 95%信用区間 (-1.61,-0.74) のかなり外側で,予測的なクロスバリデーションに基づく予測的な平均効果の 95%信用区間 (-1.98,-0.38) にはごくわずかに含まれている. 試験 25 を除く変量効果モデルによる NMA での試験間の異質性パラメータの事後中央値は 0.29 で,95%信用区間 (0.05,0.58) であり,すべてのデータ (25 試験) に基づくモデルに基づく値 (0.42) よりも小さくなった.

プログラム 2: Binomial likelihood, logit link, 予測的なクロスバリデーション, 複数群試験によるネットワークメタアナリシス (化学療法の有害事象の事例).

```
# Binomial likelihood,logit link,ネットワーメタアナリシス(3群以上の臨床試験)
# 変量効果モデル, 予測的なクロスバリデーション
                                                # *** PROGRAM STARTS
model{
for(i in 1:ns){
                                        # 試験のループ処理
                                        #3群以上の臨床試験の補正(対照群は0)
w[i,1] < 0
                                        # 対照群の治療効果は0
delta[i,1] <- 0
mu[i] \sim dnorm(0,.0001)
                                                # ベースラインの漠然事前分布
                                                # LOOP THROUGH ARMS
for (k in 1:na[i]) {
r[i,k] \sim dbin(p[i,k],n[i,k])
                                                # 治療群のループ処理
                                                # 線形予測子のモデル
   logit(p[i,k]) <- mu[i] + delta[i,k]
   rhat[i,k] <- p[i,k] * n[i,k]
                                                # 分子の期待値
   dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
                                                # 挽脱度
   + (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
                                                # 試験ごとの残差挽脱度の合計
resdev[i] <- sum(dev[i,1:na[i]])
for (k in 2:na[i]) {
                                                # 治療群のループ処理
   delta[i,k] ~ dnorm(md[i,k],taud[i,k])
                                                # 試験特有の対数オッズ比(LOR)の分布
   md[i,k] \leftarrow d[t[i,k]] - d[t[i,1]] + sw[i,k]
                                                #LOR事後分布の平均 (3群以上の試験の補正)
                                                #LOR事後分布の精度(3群以上の試験の補正)
   taud[i,k] <- tau *2*(k-1)/k
   w[i,k] \leftarrow (delta[i,k] - d[t[i,k]] + d[t[i,1]])
                                                #3群以上の臨床試験の補正
   sw[i,k] <- sum(w[i,1:k-1])/(k-1)
                                                #3群以上の臨床試験の累積補正
}
}
totresdev <- sum(resdev[])
                                                # 残差逸脱度の総合計
d[1]<-0
                                                # 参照治療1の治療効果は0
                                               # 治療効果の漠然事前分布
for (k \text{ in } 2:nt) \{ d[k] \sim dnorm(0,.0001) \}
sd \sim dunif(0,5)
                                                # 試験間標準偏差の漠然事前分布
tau <- pow(sd,-2)
                                                # 試験間精度=(1/試験間分散)
# 将来の試験に対する予測分布は多変量正規分布
delta.new[1] <- 0
                                                # 参照治療1の治療効果は0
w.new[1] <- 0
                                                # 対象治療1に対する条件付平均の補正値は0
                                                # 治療のループ処理
for (k in 2:nt) {
delta.new[k] ~ dnorm(m.new[k],tau.new[k])
                                                # 各delta.newの条件付分布
m.new[k] \leftarrow d[k] + sw.new[k]
                                                # delta.newの条件付平均
tau.new[k] <- tau *2*(k-1)/k
                                                # delta.newの条件付精度
\underline{\text{w.new[k]}} \leftarrow \text{delta.new[k]} - \text{d[k]}
                                                # 条件付平均の補正
sw.new[k] <- sum(w.new[1:k-1])/(k-1)
                                                # 条件付平均の累積補正
                                        # ベースライン (対照群) の効果を生成
p.base ~ dbeta(a,b)
                                                # 対照群のイベント数
a <- r[ns+1,1]
```

```
      b <- n[ns+1,1]-r[ns+1,1]</td>
      # 対照群の非イベント数

      for (k in 2:na[ns+1]) {
      # 治療群のループ処理

      # 新しい試験の各治療群におけるイベントの予測確率の算出
      logit(p.new[k]) <- logit(p.base) + (delta.new[t[ns+1,k]]- delta.new[t[ns+1,1]])</td>

      r.new[k] ~ dbin(p.new[k], n[ns+1,k])
      # 新しい試験の各治療群の予測イベント数の生成

      # ベイズ流のp値: 得られた値より極端な値をとる確率
      # モデル及び残りのデータを与えられたもとでの観察されたイベント数 (r[ns+1,2])との比較 p.cross[k] <- step(r[ns+1,2] - r.new[k]) - 0.5*equals(r.new[k],r[ns+1,2])</td>
      # より大きな極端な値の割合

      1
      # **** PROGRAM ENDS
```

新しい試験に対するすべての2群比較に関する対数オッズ比及びオッズ比を予測するには、以下のプログラムコードを最後の右中括弧(})の前に追加する.

```
# 3群以上の場合、すべての2群比較に対するオッズ比及び対数オッズ比の算出

for (c in 1:(nt-1)) {
  for (k in (c+1):nt) {
    lor.new[c,k] <- delta.new[k]- delta.new[c]
    or.new[c,k] <- exp(lor.new[c,k])
  }
}
```

データと連鎖ごとの初期値

```
# Data (Adverse events in Chemotherapy)
list(ns=24, nt=4)
t[,1] t[,2] na[] r[,1] r[,2] n[,1] n[,2] # ID
        2
             15
                  10
                        75
                                  # 1
  3
        2
            27
                  14
                       147
                           149
                                  #2
2
       2
            2
  3
                 5
                        25
                             46
                                  #3
(中略)
                                  # 24
  3
       2
            10 3
                      118
                           123
1
   3
             78
                        465
                             463
                                  # 25
1
END
# Initial values
# Initial values for delta and other variables can be generated by WinBUGS.
#chain 3
list(d=c(\ NA,\ 2,3,-3),\ sd=2,\ mu=c(-3,5,-1,-3,7,\ -3,-4,-3,-3,0,\ -3,5,-1,-3,7,\ -3,-4,-3,-3,0,\ -3,-3,0,3),\ p.base=.8)
```

NMAの文脈では、外れ値の検出に対するクロスバリデーションは、Node-splitting法[34]のような不一致性の点検法に密接に関係している。Node-splitting法では、ある特定の対比に関する「直接的な」エビデンスが、残りのネットワークから推定された治療効果の推定値と乖離する場合は、残りのネットワークから予測された治療効果と比較される。実際に、Node-splitting法は、ただ1つの試験ではなく複数試験のサブセットが、オリジナルの解析から取り除かれるクロスバリデーションに類似している。

しかしながら、本手法との1つの重要な違いは、クロスバリデーションが本質的に「外れ値」を検出す

るための手法であるのに対し(補足:クロスバリデーション自体は、モデルの妥当性やバイアス補正などにも広く使われている)、NMAでの「直接的」及び「間接的」エビデンスの間の不一致性は、期待される(すなわち平均での)効果の不一致を指している。この理由から、TSD4[7](4章)に示すように、不一致性の点検に用いるNode-splitting法は、平均効果の予測分布ではなく、事後分布に基づいている。1つの辺(edge)が1つの試験で構成される三角ネットワークでは、Node-splitting法が期待効果での不一致性を示すが、一方クロスバリデーションではその1試験を外れ値として検出できないような状況が頻繁に起こり得る。そのような結果は決して矛盾しているわけではない。すなわち、マグネシウムの例でのISIS-4試験は、残りの試験のエビデンスに基づいた期待効果(事後平均)からは非常に顕著に外れているものの、予測分布が考慮されるとき、外れ値にはならない。しかしながら、化学療法の有害事象の例では、例えばABとACの平均効果が(1つの試験に基づくものではあるが)BCの平均効果と整合しない等のエビデンスを検討しているかどうか、又はそれらが1つのBC試験が、エビデンス統合の文脈において、「外れ値」であると懸念されるかどうかを、研究者は明らかにする必要があることを示している。もちろん、技術的には、治療効果の予測分布に基づいて、不一致性の点検ができない理由はなく、推定がネットワークメタアナリシス(NMA)での予測治療効果に基づくのであれば、それが望ましい。

3.7 サブグループ, メタ回帰, ベースラインリスクへの調整

3.7.1 メタ回帰の導入: サブグループ効果

無作為化比較試験の治療効果に関連して、サブグループ効果は、治療との交互作用をもつカテゴリカルな試験レベルの共変量として理解でき、3.2.1節のシナリオA1に対応する.この場合の仮説は、治療効果の大きさが、例えば、男性、女性の患者で異なる、又は治療効果の大きさが年齢グループ、前治療などに依存している等になる.そのようなデータを解析する最も簡単な方法は、サブグループ毎に別々の解析を実施した後、相対治療効果の推定値を推定するものである.しかしながら、そのアプローチは2つの欠点がある.1つ目の問題は、モデルが変量効果を有する場合、サブグループ毎に別々の解析を実施することは、試験間変動に対する異なる推定値が得られることを意味する.試験間変動を推定するための十分なデータがあることは稀であるため、全てのサブグループに対して治療効果が同じであると仮定する方が、より理にかなっているかもしれない.2番目の問題は、サブグループ毎に別々の解析を実行することは、サブグループ間で治療効果が等しいとする帰無仮説を棄却するために必要な交互作用の検定を直接的には実施できないことである.サブグループ毎に別々に解析を実施する方法の代替法は、治療効果に導入される、共通の試験間の異質性パラメータ、及び交互作用項βを伴った1つの統合解析である.TSD2[8]で紹介した、別々の2群比較のメタアナリシスに対する変量効果モデルは、以下である.

$$\theta_{ik} = \mu_i + \delta_{i,1k} I_{\{k \neq 1\}}$$

ここに、 θ_{ik} は試験i、群kでの線形予測子(例えば、対数オッズ)であり、 μ_i は試験iでの試験特有のベースライン効果であり、関連しない局外母数(nuisance parameter)として見なされ、 $\delta_{i,1k}$ は試験iの対照群(群1)に対する群kの試験特有の治療効果で、Iは式(1)で定義される.

変量効果を伴うメタ回帰モデルは以下となり,

$$\theta_{ik} = \mu_i + \left(\delta_{i,1k} + \beta x_i\right) I_{\{k \neq 1\}} \tag{4}$$

ここで、 x_i は試験iの試験レベルの共変量で、サブグループ、連続値の共変量、又はベースラインリスクを表せる。また式(4)は、次のように書き直すことができる。

$$\theta_{i1} = \mu_i$$

$$\theta_{i2} = \mu_i + \delta_{i,12} + \beta x_i$$

治療法と共変量との交互作用(δ 及び β)は治療群のみで考慮され、対照群では考慮されないことに注意する. 変量効果モデルでは、試験特有の対数オッズ比は共通の分布 $\delta_{i,12}\sim N(d,\sigma^2)$ から得られるとする. 固定効果モデルでは、式(4)を $\theta_{ik}=\mu_i+(d+\beta x_i)I_{\{k\neq 1\}}$ に置き換える. ベイズの枠組みでは、d、 β 、 σ は独立な(無情報の)事前分布が与えられる。例えば、d, $\beta\sim N(0,100^2)$ 、及び $\sigma\sim Uniform(0,5)$ である.

以下に、サブグループ毎に別々の解析から得られる結果と、サブグループの交互作用解析から得られた結果を対比した事例を与えた.

理想的には、可能であれば臨床専門家から引き出された情報に基づく事前分布を用いて、「統計的に有意」であるかのサブグループ項を含めたい。しかしながら、NICE Methods Guideでは、サブグループ効果が費用対効果モデルで考慮される場合は、事前分布の正当性を有すると共に、サブグループ効果が統計的に頑健(ロバスト)であるべきであると提言している。実際、統計的に弱い交互作用に基づいて、ある治療法が採択又は棄却されるべきとの議論を続けることは難しいだろう。

2 群比較のメタアナリシスでのサブグループ: スタチンの事例

コレステロール降下効果をプラセボ又は通常治療とスタチンを比較した 19 試験のメタアナリシス[35] には、1 次予防(心疾患の既往なし)が目的の試験が幾つか含まれ、残りは2次予防(心疾患の既往あり)が目的の試験が含まれていた。サブグループの指標は試験レベルの共変量であることに留意する。興味のあるアウトカム指標は、全死因による死亡率で、

表 3-3にデータを示す. 潜在的な効果修飾因子 (effect-modifier) である, 1次予防 対 2次予防は,全 てのデータの2群比較のメタアナリシスにて,サブグループを考慮すること,又は2つのタイプの試験ごとに,別々のメタアナリシスを実施することができる.

表 3-3 心疾患の既往あり/なし患者におけるプラセボに対するスタチンのメタアナリシス[35]: 19 つの RCT における対照群及びスタチン群でのすべての原因による死亡数

	Placebo/U	Usual care	Sta	atin	
Trial	# of deaths	# of patients	# of deaths	# of patients	Type of prevention
ID	r_{i1}	n_{i1}	r_{i2}	n_{i2}	x_i
1	256	2223	182	2221	Secondary
2	4	125	1	129	Secondary
3	0	52	1	94	Secondary
4	2	166	2	165	Secondary
5	77	3301	80	3304	Primary
6	3	1663	33	6582	Primary
7	8	459	1	460	Secondary
8	3	155	3	145	Secondary
9	0	42	1	83	Secondary
10	4	223	3	224	Primary
11	633	4520	498	4512	Secondary
12	1	124	2	123	Secondary
13	11	188	4	193	Secondary
14	5	78	4	79	Secondary
15	6	202	4	206	Secondary
16	3	532	0	530	Primary
17	4	178	2	187	Secondary
18	1	201	3	203	Secondary
19	135	3293	106	3305	Primary

[出典: NICE DSU TSD3 (http://nicedsu.org.uk/technical-support-documents/)]

試験i、群kでの死亡数 r_{ik} は、二項尤度 r_{ik} ~Binomial(p_{ik} , n_{ik})、i=1,...,19; k=1,2に従うと仮定する. x_i を以下のような試験レベルのサブグループ指標として定義する.

$$x_i = \begin{cases} 0 & 試験iが1次予防患者の試験 \\ 1 & 試験iが2次予防患者の試験 \end{cases}$$

ここでの交互作用のモデルは式(4)で与えられ、ここに $\theta_{ik} = \text{logit}(p_{ik})$ は線形予測子(TSD2[8]参照)である。本設定では、 μ_i は対照治療(すなわち、治療1)のアウトカムの対数オッズを表し、 $\delta_{i,12}$ は1次予防の試験での対照群に対する治療群での試験特有の成功の対数オッズ比である。

2つの別々の固定効果モデル又は変量効果モデルの当てはめるためのWinBUGSのコードはTSD2 (Program 1(a)~1(d)) に与えられている。サブグループの交互作用項を含めた1つの解析についてのWinBUGSコードを以下に与える。本事例には2つの治療法しか含まれていないが,以下に示すプログラムコードは,複数の治療法,及び複数群の試験を含むサブグループ解析にも用いることができる。メタ回帰におけるサブグループの変量効果モデルに対するWinBUGSプログラムコードはプログラム3(a)に,固定効果モデルに対するプログラムコードはプログラム3(b) に与えられている。

プログラム 3 (a): Binomial likelihood, logit link, 変量効果, サブグループによるメタ回帰分析 (スタチンの事例)

```
# Binomial likelihood, logit link, サブグループ
#3群以上の試験に対する変量効果モデル
                                               # *** PROGRAM STARTS
model{
for(i in 1:ns){
                                        # 試験のループ処理
w[i,1] < 0
                                        #3群以上の臨床試験の補正(対照群は0)
delta[i,1] <- 0
                                        # 対照群の治療効果は0
mu[i] \sim dnorm(0,.0001)
                                               # ベースラインに対する漠然事前分布
                                               # 治療群のループ処理
for (k in 1:na[i]) {
   r[i,k] \sim dbin(p[i,k],n[i,k])
                                               # 二項尤度
   # 線形予測子に対するモデル,1つ目の群に対する共変量の効果
   logit(p[i,k]) \leftarrow mu[i] + delta[i,k] + (beta[t[i,k]]-beta[t[i,1]]) * x[i]
   rhat[i,k] <- p[i,k] * n[i,k]
                                               # 分子の期待値
   dev[i,k] \leftarrow 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
                                               # 逸脱度
   + (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
                                               # 試験ごとの残差逸脱度の合計
resdev[i] <- sum(dev[i,1:na[i]])
for (k in 2:na[i]) {
                                               # 治療群のループ処理
                                               # 試験特有の対数オッズ比(LOR)の分布
delta[i,k] ~ dnorm(md[i,k],taud[i,k])
md[i,k] \leftarrow d[t[i,k]] - d[t[i,1]] + sw[i,k]
                                               #LOR事後分布の平均(3群以上の試験の補正)
taud[i,k] <- tau *2*(k-1)/k
                                               #LOR事後分布の精度(3群以上の試験の補正)
w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]])
                                               #3群以上の臨床試験の補正
sw[i,k] <- sum(w[i,1:k-1])/(k-1)
                                               #3群以上の臨床試験の累積補正
}
}
totresdev <- sum(resdev[])
                                               # 残差逸脱度の総合計
d[1]<-0
                                               # 参照治療1の治療効果は0
beta[1] <- 0
                                        # 参照治療1の共変量効果は0
for (k in 2:nt){
                                        # 治療のループ処理
d[k] \sim dnorm(0,.0001)
                                        # 治療効果の漠然事前分布
beta[k] <- B
                                        # 共通の共変量効果
}
B \sim dnorm(0,.0001)
                                        # 共変量効果の漠然事前分布
sd \sim dunif(0,5)
                                               # 試験間標準偏差の漠然事前分布
tau <- pow(sd,-2)
                                               # 試験間精度=(1/試験間分散)
                                        # *** PROGRAM ENDS
```

関心のある他のパラメータに対する事後分布からの要約を得るには、d、B及びsdをモニターする必要がある。モデルの適合度とモデルの比較を検討するために必要なパラメータの事後平均値を得るには、dev、totresdev及び(WinBUGS DIC ツールから得られる)DIC をモニターする必要がある。

群1の治療法に対する絶対治療効果に関する追加情報が与えられれば、与えられた共変量の値に対する、すべての2群の対数オッズ比及びオッズ比を推定し、絶対効果の推定値を生成するため、最後の右中括弧の前にプログラムコードを追加する。結果から他の要約を計算すること、及びその要約を他の尺度への変換に関するより詳細な内容は、TSD2[8]の付録を参照のこと。

```
#(データとして与えられた)長さnzの共変量ベクトルzに関して、
# すべてのオッズ比, 対数オッズ比, 絶対効果を算出するコード
for (k in 1:nt){
for (j in 1:nz) { dz[j,k] <- d[k] + (beta[k]-beta[1])*z[j] } # 共変量がz[j]の時の治療効果
# すべての2群比較に対するオッズ比及び対数オッズ比
for (c in 1:(nt-1)) {
 for (k in (c+1):nt) {
# 共変量が0の場合
    or[c,k] \leftarrow exp(d[k] - d[c])
    lor[c,k] \leftarrow (d[k]-d[c])
# 共変量がz[j]の場合
     for (j in 1:nz) {
       orz[j,c,k] \leftarrow exp(dz[j,k] - dz[j,c])
       lorz[j,c,k] \leftarrow (dz[j,k]-dz[j,c])
     }
    }
}
# 参照治療1に対する平均効果 (meanA) , 精度 (=1/分散) (precA) 及び共変量 z[j] を与えたもとで,
# 治療効果 T[k] の推定値を算出
A ~ dnorm(meanA,precA)
for (k in 1:nt) {
for (j in 1:nz){
 logit(T[j,k]) \leftarrow A + d[k] + (beta[k]-beta[1]) * z[j]
 }
}
```

2 つのサブグループを持つメタ回帰分析の場合、ベクトルz が list(z=c(1), nz=1)として、リストデータステートメントに追加される.

データ構造は TSD2[8]に示されている構造と同じだが、各試験の共変量の値(0 又は 1)を表す追加の列 x []がある.残りの変数は、治療法の数 nt、試験数 ns、r[,1]及び n[,1]は最初の治療の分子及び分母、r[,2]及び n[,2]は 2 番目の治療の分子及び分母、t[,1]及び t[,2]は最初と 2 番目の治療法(番号)の識別子、na []は各試験での群数である.

データと連鎖ごとの初期値

```
# Data (Statins example)
list(ns=19, nt=2)
t[,1] t[,2]
         na[]
                r[,1]
                     n[,1]
                            r[,2]
                                 n[,2] x[]
                                              # ID name
                256
                      2223
                           182
                                  2221
                                              #14S
          2
                                        1
    2
          2
                4
                      125
                           1
                                  129
                                        1
                                              # 2 Bestehorn
1
         2
1
    2
               0
                      52
                            1
                                  94
                                        1
                                              #3 Brown
 (中略)
   2
          2
               1
                      201
                                  203
                                        1
                                              # 19 Weintraub
          2
               135
                      3293 106
                                  3305 0
                                              # 20 Wscotland
    2
1
END
# Initial values
# Initial values for delta can be generated by WinBUGS.
#chain 2
#chain 3
list(d=c( NA, 2), mu=c(-3,5,-1,-3,7, -3,-4,-3,-3,0, 5,0,-2,-5,1, -2,5,3,0), B=1.5, sd=0.5)
```

プログラム 3 (b): Binomial likelihood, logit link, 固定効果, サブグループによるメタ回帰分析 (スタチンの事例)

```
# Binomial likelihood, logit link, サブグループ
#1つの共変量を含む固定効果モデル
model{
for(i in 1:ns){
mu[i] \sim dnorm(0,.0001)
for (k in 1:na[i]) {
    r[i,k] \sim dbin(p[i,k],n[i,k])
    logit(p[i,k]) <- mu[i] + d[t[i,k]] - d[t[i,1]] + (beta[t[i,k]] - beta[t[i,1]]) * x[i]
    rhat[i,k] \leftarrow p[i,k] * n[i,k]
    dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
    + \; (n[i,k]\text{-}r[i,k]) \;^* \; (log(n[i,k]\text{-}r[i,k]) \;\text{-}\; log(n[i,k]\text{-}rhat[i,k])))
}
resdev[i] <- sum(dev[i,1:na[i]])
}
totresdev <- sum(resdev[])
d[1] < 0
beta[1] <- 0
for (k in 2:nt){
d[k] \sim dnorm(0,.0001)
beta[k] <- B
B \sim dnorm(0,.0001)
```

2つの別々の解析と、固定効果モデル及び変量効果モデルに交互作用モデルを用いた1つの解析結果 (TSD2[8]で紹介したモデル適合度に関する統計量を含めた)を表 3-4に示す。固定効果モデルでは、別々の解析では10,000回のBurn-in後に収束し(1つのモデルでの解析では20,000回のBurn-in後)、結果は 3つの独立連鎖からの50,000サンプルに基づいている。変量効果モデルでは、別々の解析では40,000回の

Burn-in, 1つのモデルでの解析では50,000回のBurn-inが利用された。結果は3つの独立連鎖からの100,000 サンプルに基づいている。固定効果モデルでは、別々の解析と1つの解析で治療効果は正確に同じ結果になる。一方で、変量効果モデルでは、2つのサブグループで共通の分散を仮定するため、治療効果が必ずしも全く同じとはならない。特に1次予防のサブグループでは、分散パラメータを得るために利用可能なエビデンスがほとんどなく(5試験)、別々の変量効果メタアナリシスでの全ての推定値において、非常に広い信用区間(CrI)となっており、1つモデルでの解析により、推定値はより正確となる(信用区間が狭くなる)。しかしながら、ベイズの枠組みでは、1つのモデルでの解析のみが交互作用項 β の直接的な検定を提供しており、 β の点推定値は負であることから、2次予防でスタチンがより効果的であることを示唆していたものの、固定効果及び変量効果の両方のモデルで、 β の95%信用区間(CrI)は交互作用がない可能性を示していた(β の95%信用区間が0を含んでいた)。

表 3-4 事後分布の要約:対数オッズ比 (LOR) , オッズ比 (OR) 及び試験間の異質性 (バラツキ; σ) に対する事後平均値 (LOR, OR) /事後中央値 (σ) , 標準偏差 (σ) , 交互作用項 (σ) 及び 95%信用区間 (CrI) モデル適合度: 残差逸脱度の事後平均 (resdev) , 有効パラメータ数(σ)及び DIC

	Fixed	effects	Randor	n effects	
	Primary Prevention	Secondary Prevention	Primary Prevention	Secondary Prevention	
	Separate	e analyses	Separate	e analyses	
	mean (sd) [95%Crl]	mean (sd) [95%Crl]	mean/median (sd) [95% Crl]	mean/median (sd) [95%Crl]	
LOR	-0.11 (0.1)	-0.31 (0.05)	-0.18 (0.74)	-0.36 (0.16)	
	[-0.30, 0.09]	[-0.42, -0.21]	[-2.01, 1.12]	[-0.72, -0.06]	
OR	0.9 (0.09)	0.73 (0.04)	1.12 (3.65)	0.71 (0.11)	
	[0.74, 1.09]	[0.66, 0.81]	[0.13, 3.07]	[0.49, 0.94]	
σ	-	-	0.79 (0.98)	0.16 (0.23)	
			[0.06, 3.90]	[0.01, 0.86]	
resdev	16.9 [†]	29.0 [‡]	11.9 [†]	28.3 [‡]	
pD	6	15	9.3	16.8	
DIC	22.9	44	21.1	45.1	
		th interaction term, ubgroup	Single analysis with interaction term, β , for subgroup		
	mean (sd) [95%Crl]	mean (sd) [95%Crl]	mean/median (sd) [95%Crl]	mean/median (sd) [95%Crl]	
β	-0.21 (0.11)		-0.29 (0.26)		
LOR	[-0.42, 0.01] -0.11 (0.1) [-0.30, 0.09]	-0.31 (0.05) [-0.42, -0.21]	[-0.86, 0.20] -0.07 (0.20) [-0.48, 0.36]	-0.36 (0.16) [-0.72, -0.07]	
OR	0.9 (0.09)	0.73 (0.04)	0.95 (0.21)	0.7 (0.11)	
	[0.74, 1.09]	[0.66, 0.81]	[0.62, 1.43]	[0.49, 0.94]	
σ	-	-	0.19 (0.20) [0.01, 0.76]		
resdev*	45.9		42.6		
pD	21.0		25.0		
DIC	66.9		67.6		

^{†10 (}データポイント) と比較, [‡]28 (データポイント) と比較, ^{*}38 (データポイント) と比較 固定効果モデル: 別々の解析では 10,000 回の Burn-in 後に収束し (1 つのモデルでの解析では 20,000 回の burm-in 後), 結果は 3 つの独立連鎖からの 50,000 サンプルに基づく(WinBUGS)

変量効果モデル: 別々の解析では 40,000 回の Burn-in, 1 つのモデルでの解析では 50,000 回の Burn-in が利用された. 結果は 3 つの独立連鎖からの 100,000 サンプルに基づく(WinBUGS)

注意:変量効果モデルでは、対数オッズ比とオッズ比が平均値又は中央値で記載されている.

[出典: NICE DSU TSD3 http://nicedsu.org.uk/technical-support-documents/technical-support-documents/)]

必ずしも容易ではないが、この考え方は2値から複数カテゴリの効果修飾因子に自然に拡張できる. 例えば、患者が軽度、中程度、重度にカテゴリ化された試験では、2つの交互作用項、すなわち1つ目は軽度と比べた中程度の交互作用、2つ目は軽度と比べた重度の交互作用を導入することができる. その代わりに、疾患の重症度は連続値の共変量として(3.7.3節参照)、又はベースラインリスクに対する回帰として検討することができる(3.7.4節参照). 更なる変法は交互作用項を変量効果として導入することである. 意思決定での適用はおそらく稀であるが、そのようなモデルは国や地域間での治療効果における変動の解析で価値があるかもしれない. その場合、地域内で十分な数の試験が統合のために利用可能であることを仮定している. この場合、異なる交互作用項が各地域で提案され、これらはある平均と地域間の分散を有する共通の分布からランダムに抽出されると考える. S個の試験のメタアナリシスでは、変量効果の交互作用モデルは以下であり、

$$\theta_{ik} = \mu_i + \left(\delta_{i,1k} + \beta_i x_i\right) I_{\{k \neq 1\}}$$

試験i (i=1,...,S)が地域j, k=1,2で実施されたとき $\beta_i = B_i$ となる.

$$B_i \sim N(b, \tau_b^2)$$

ここに、 B_j は地域特有の交互作用を示し、bは地域間を通じた交互作用の平均効果を表し、 au_b^2 は地域間の変動を表す.

3.7.2 NMA における交互作用モデルの範囲とその解釈

原則として、前節と同じ考え方が多群試験のネットワークによる統合へ適用される.しかしながら、非常に多くのモデルが提案されており、それぞれかなり異なる意味合いを持つ.以下では、利用可能なモデルの範囲を提示し、それらの解釈を議論する.サブグループ効果に関してモデルの範囲を検討するが、サブグループの交互作用モデルは、連続量を共変量としたメタ回帰(3.7.3 節参照)、又はベースラインリスクでのメタ回帰(3.7.4 節参照)と構造的に同じであり、連続量のケースでも正確に同じ範囲のモデルが検討できる.限られた交互作用モデルのクラスのみが、実用的な意思決定にて、有用な解釈を与える.多群試験のメタ回帰モデルへの適用に関して、一般的な3つのアプローチを提示する.1つ目は治療法間で関係がなく(交換可能性がなく)、サブグループと治療法との交互作用項を検討するもの、2つ目は交換可能性があり、治療法間で関係するサブグループと治療法との交互作用項を検討するもの、3つ目は全ての治療法との1つのサブグループと治療法との交互作用を検討するものである.

ここで、交互作用の推定において、「関係がある」とは治療法間での交換可能性を考える場合である。 それに対し「関係がない」とは、交互作用の交換可能性を考えない場合である。交換可能性を仮定する と、各交互作用は共通の分布からのサンプリングとなり、その共通の分布もデータから推定するため、 間接的に影響を与え合う(関係する)ことになる。

1 次予防 vs 2 次予防を例として,S 個の複数の治療法 $T_1, T_2, ..., T_s$ を比較する場合,試験間の共変量として 2 値の場合を考える.TSD2[8]での一致性のモデルに対するアプローチに従うと,治療法 1 に対する各治療法の相対治療効果である s-1 個の基本パラメータ $d_{12}, d_{13}, ..., d_{1s}$ を考えることになる.これま

でと同様に、治療法 1 はプラセボ又は標準治療であると仮定し、NMA の参照治療とする。残りの(s-1)(s-2)/2 個の治療法の対比は、一致性の式を用いて、これらのパラメータを用いて表現される。例えば、治療法 3 と治療法 4 との相対効果は、 $d_{34}=d_{14}-d_{13}$ と書ける。1 で詳述するように、一連の固定効果による交互作用モデルを設定できる。これらのモデルは、治療効果の試験間変動を考慮するモデルに、容易に拡張可能である。事例は 3.7.3 節及び 3.7.4 節に与える。

Box 1 交互作用モデルの範囲

1. 治療法ごとの独立な交互作用(Independent, treatment-specific interactions)

この場合,例えば1次予防又は2次予防と治療法との間に交互作用がある場合,これらの交互作用を治療ごとに異なるものと考える場合である.これをモデル化するには,基本治療法と同数の交互作用項の数を導入する.例えば, β_{12} , β_{13} , ..., β_{1s} となる.

これらの各項は、治療法1と治療法2、3、…、sとの比較における1次予防に比した2次予防の追加効果(交互作用)を表している.これらの項は、主効果 $d_{12},d_{13},\dots,d_{1s}$ 、すなわち1次予防患者での治療効果であるが、これらと正確に平行である.例えば、治療法3と治療法4を比較する試験での主効果の検討と同様に、交互作用項は治療法1への相対治療効果の交互作用項の差 $\beta_{34}=\beta_{14}-\beta_{13}$ となる.TSD2[8] の表記にしたがうと、線形予測子の固定効果モデルは次のようになる.

$$\theta_{ik} = \mu_i + \left(d_{t_{i1},t_{ik}} + \beta_{t_{i1},t_{ik}} x_i\right) I_{\{k \neq 1\}} = \mu_i + \left(d_{1t_{ik}} - d_{1t_{i1}} + \left(\beta_{1t_{ik}} - \beta_{1t_{i1}}\right) x_i\right) I_{\{k \neq 1\}}$$
 (5)

 t_{ik} は試験 i の群 k の治療法を表し、 x_i は共変量又はサブグループの指示変数、I は式(1)で定義した。本モデルでは、 $d_{11}=\beta_{11}=0$ とする。残りの交互作用項は無関係(unrelated)であり(互いに影響されない)、ベイズ解析で関係のない(互いに影響を与え合わない)漠然事前分布を与えられる。したがって、2 次予防での相対治療効果は $d_{12}+\beta_{12}$, $d_{13}+\beta_{13}$, ..., $d_{1s}+\beta_{1s}$ である。実際、本モデルの解釈は以下のようになる。1 次予防患者でのs 個の各治療法の相対治療効果は、2 次予防患者での対応する相対治療効果とは、完全に無関係となる。実際に 1 次予防と 2 次予防を別々に解析しようとする場合、交互作用項の検定することがより難しくなり、3.7.1 節で示したように、変量効果モデルで分散を共有する(shared variance terms)こともできなくなる。

2. 治療法間で交換可能な交互作用 (Exchangeable, related, treatment-specific interactions)

本モデルは上記 1 のモデルと同じ構造で、パラメータ数も同じである。唯一の違いは(s-1)個の「基本的な」交互作用項は、治療間で独立な漠然事前分布(unrelated vague priors)を与えるのではなく、共通の平均と分散をもつ確率分布 $\beta_{1k} \sim N(b,\tau^2)$ 、k=2,...,s から抽出されることである。交互作用項が薬剤間でどの程度類似しているか(又は異なっているか)に対して制限を置く情報のある事前分布(informative prior)を使うことが可能であるが、交互作用の平均及び分散はデータから推定される。

3. 全ての治療法で同じ交互作用効果 (補足:全ての治療法で1つの交互作用項) (The same interaction effect for all treatments)

最後のモデルでは、1つの交互作用項 b を治療法 1 に対する全ての治療法の治療効果に対して適用する。 つまり全ての治療法 k=2,...,s に対し $\beta_{1k}=b$ とする。したがって、治療法 1 に対する相対治療効果 $d_{12},d_{13},...,d_{1s}$ は全て 2 次予防での治療効果 $d_{12}+b,d_{13}+b,...,d_{1s}+b$ より、同じb だけ高く又は低くなる。ただし、交互作用項が相殺されるため、1 次予防及び 2 次予防患者での治療法同士(2,3,...,s)の相対治療効果は正確に同じになる。これは治療法 1 を参照に選ぶことが重要であり、また共変量を含むモデルの結果はその選択に影響を受ける(sensitive)ことを意味する。モデルをコーディングする際、パラメータの解釈を意識すべきである。例えば、2 次予防患者での治療法 3 に対する治療法 4 の相対治療効果は $d_{14}+b$ - $(d_{13}+b)$ = d_{14} - d_{13} であり、1 次予防患者での相対治療効果と同じとなる。

効果修飾因子(effect-modification)を許容するモデルの場合,モデル構築での一連の選択肢がある. モデル選択に影響しうる 1 つの要因は,利用可能なデータ量である.固定効果モデルで考える場合,関係のない交互作用モデル(unrelated interaction model; Box 1 のモデル 1)は,各サブグループで少なくとも s-1 個の試験を有するように,全ての治療法を含む 2 つの連結したネットワーク(各サブグループが 1 つのネットワーク)が必要となる.変量効果モデルでは共通の試験間分散を推定するため,より多くのデータが必要となる.

交換可能な交互作用モデル(Exchangeable interaction model; Box 1でのモデル2)では少ないデータでも推定可能かもしれない. ただし、本モデルを用いるには、交換可能性の明確な根拠が必要となる. 1つの理論的根拠は、同じクラスに含まれる治療法に対し交換可能な交互作用を仮定し、異なる共変量との交互作用を許容するものである. したがって、治療法2,...,sの幾つかは1つの「クラス」に属し、治療法1は標準治療又はプラセボとする. 例えば、1つの交換可能な交互作用項としては、心房細動(AF)に対する、プラセボに対するアスピリンを基礎とした治療法のセット、別のセットとしては、プラセボに対するワルファリンを基礎とした治療法である[36].

交換可能な交互作用は一見すると魅了的なアプローチではあるが、十分なデータがあっても、臨床現場及び意思決定で用いることは、反直感的であり、反論が難しい勧告をされ得るという困難がある。本モデルの主張は、そのクラス内の治療法の相対治療効果に実際に違いがあるということである。治療の推奨に対して、モデル1又はモデル2に基づく場合、増分費用効果分析を厳格に適用すると、異なるサブグループで異なる治療法が推奨され得る。交互作用が異なるという仮説が、統計的に頑健(ロバスト)であることが示されていない限り、その結論は逆の結果(異なるサブグループで異なる治療法を推奨され得ない)と見なされる可能性がある(補足:ある新薬が、「少なくともあるサブグループでは有効/有用である」と主張とするためには、頑健な解析結果を示さない限り、認められる可能性は小さいということを示唆している)。

これらの理由から、本TSDでは、Box 1で述べた3つの一般的なモデルのうち、最後のモデルのみを検討する. すなわち、治療法1(参照治療)に対する全ての治療法を通じて、同一の交互作用を仮定したモ

デルである. 3.7.3節で事例を紹介する. 関係のない交互作用又は交換可能な交互作用に対する代替モデルを排除するわけではない. つまり, (これらのモデルを用いて) 探索的解析又は仮説形成の検討にて, 確かに役割を果たせる. 事例及びコーディング法については文献を参照されたい[36][37].

よりもっと限定されたモデルを提案することが妥当な場合がある。クラス内の全ての実薬の治療法に対し、1つの交互作用項を用いるというより、クラスに関わらず全ての実薬の治療法に対し、単純に1つの交互作用項を用いることができる。例えば、幾つかの治療法は効果的であり、実質的に有害な症状を排除できる場合である。すなわち、より深刻に影響を受ける患者では、改善の程度が必然的により大きくなるため、重症度と治療効果との「交互作用」が存在することは、ほとんど避けられない。

潜在的な事例としては、炎症性関節炎のための生物(学的)療法の異なるクラス、又は恐らく疼痛緩和のための特定の治療法などである.これらの場合、「交互作用」は、治療の薬理学的効果というよりは、測定尺度の特性を反映しているのかもしれない。これまで通り、モデル式への臨床的及び科学的な知見を反映させることは重要である.

3.7.3 連続値の共変量でのメタ回帰

連続量の共変量を扱うとき、MCMC法での連鎖の混交時間(mixing time)を改善するには、共変量の中心的な値を解析で用いるべきである。これは共変量の各値 x_i から平均値 \bar{x} を差し引くことで可能となる。単純な2群比較のメタアナリシスの場合、式(4)は以下のようになる。

$$\theta_{ik} = \mu_i + (\delta_{ik} + \beta(x_i - \bar{x}))I_{\{k \neq 1\}}$$
 (6)

治療効果は共変量の平均値において推定され、治療効果を共変量の任意の値において推定するためには、非中心化で変換することができる。つまり、共変量の値、z、における平均治療効果は、 $d-\beta(\bar{x}-z)$ である(補足:共変量の平均値(\bar{x})における治療効果はdであるが、連続値の共変量と治療効果との線形関係を利用すれば変換可能)。NMAでは、式(5)のモデルを同様に中心化できる。

連続値の共変量での2群比較のメタ回帰: BCG ワクチンの例

結核 (TB) 予防のBCGワクチンの有効性を評価する試験のメタアナリシスでは, 試験が実施された絶対緯度又は赤道からの距離が, ワクチンの有効性に影響を及ぼす可能性を示唆された[38]. この例は, 3.2.1節のシナリオB1に対応する. 利用可能なデータ (表 3-5) は, 試験が実施された絶対緯度に加えて, ワクチン接種及び非接種の人数, 試験の追跡期間で結核と診断された各グループでの患者数であった.

表 3-5から求めた未調整のオッズ比を, 赤道からの距離に対してプロット (対数尺度) した (図 3-6). 規模が大きくて, 精度がより高い試験は, より直径が大きい円になるように, 各試験のプロットの円 (〇) の大きさは, 試験の精度に比例している. 線形関係 (対数オッズ比の尺度) に従って, 緯度が異なると, ワクチンの効果も異なるように見える.

表 3-5 BCG の例: ワクチン接種群及びワクチン非接種群での総患者数n のうち結核 (TB) と診断された患者数r, 及び試験実施時での絶対緯度x

	Not vac	Not vaccinated Vaccinated				
Trial number	number diagnosed with TB r_{il}	total number of patients n_{il}	number diagnosed with TB r_{l2}	total number of patients n_{i2}	Absolute degrees latitude x_i	
1	11	139	4	123	44	
2	29	303	6	306	55	
3	11	220	3	231	42	
4	248	12867	62	13598	52	
5	47	5808	33	5069	13	
6	372	1451	180	1541	44	
7	10	629	8	2545	19	
8	499	88391	505	88391	13	
9	45	7277	29	7499	27	
10	65	1665	17	1716	42	
11	141	27338	186	634	18	
12	3	2341	5	2498	33	
13	29	17854	27	16913	33	

[出典: NICE DSU TSD3 (http://nicedsu.org.uk/technical-support-documents/technical-support-documents/)]

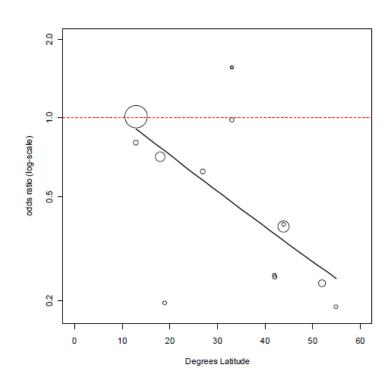


図 3-6 結核予防の BCG ワクチン: 対数尺度での赤道からの絶対緯度に対する未調整の対数オッズ比のプロット. 円の大きさは研究の精度に比例し、水平線(破線)は治療効果がない閾値を表し、実線は変量効果の交互作用モデルにより推定された回帰直線である. オッズ比が1より小さい場合、ワクチンが有効であることを示す.

[出典: NICE DSU TSD3 (http://nicedsu.org.uk/technical-support-documents/technical-support-documents/)]

試験 i の群 k で結核(TB)と診断された患者数を二項尤度 $r_{ik} \sim Binomial(p_{ik}, n_{ik})$ と仮定する. 連続値の共変量 x_i は絶対緯度を表し、式(6)のメタ回帰モデルを、共変量の平均値 $\bar{x} = 33.46$ °のもと、固定効果と変量効果モデルの両方に当てはめた. 治療効果は共変量の平均値に対する対数オッズ比が得られる.

WinBUGS のコードを以下に示す.連続値の共変量を含むメタ回帰分析による変量効果モデルのWinBUGS コードはプログラム 4 (a) に、固定効果モデルのプログラムコードはプログラム 4 (b) にある.このコードは、複数の治療法があり、複数群の試験のネットワークにも用いることができる(次節の Certolizumab の例参照).

プログラム 4 (a): Binomial likelihood, logit link, 変量効果,連続値の共変量によるメタ回帰分析 (BCG ワクチンの事例)

```
# Binomial likelihood, logit link, 連続値の共変量
#3群以上の試験に対する変量効果モデル
                                               # *** PROGRAM STARTS
model{
for(i in 1:ns){
                                        # 試験のループ処理
w[i,1] < 0
                                        #3群以上の臨床試験の補正(対照群は0)
delta[i,1] <- 0
                                        # 対照群の治療効果は0
                                               # ベースラインの漠然事前分布
mu[i] \sim dnorm(0,.0001)
for (k in 1:na[i]) {
                                               # 治療群のループ処理
   r[i,k] \sim dbin(p[i,k],n[i,k])
                                               # 二項尤度
   # 線形予測子に対するモデル、1つ目の群に対する(センタリングした)共変量の効果
   logit(p[i,k]) \leftarrow mu[i] + delta[i,k] + (beta[t[i,k]]-beta[t[i,1]]) * (x[i]-mx)
   rhat[i,k] \leftarrow p[i,k] * n[i,k]
                                               # 分子の期待値
   dev[i,k] \leftarrow 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
                                               # 逸脱度
   + (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
resdev[i] <- sum(dev[i,1:na[i]])
                                               # 試験ごとの残差逸脱度の合計
for (k in 2:na[i]) {
                                               # 治療群のループ処理
   delta[i,k] ~ dnorm(md[i,k],taud[i,k])
                                               # 試験特有の対数オッズ比(LOR)の分布
   md[i,k] \leftarrow d[t[i,k]] - d[t[i,1]] + sw[i,k]
                                               #LOR事後分布の平均(3群以上の試験の補正)
                                               #LOR事後分布の精度 (3群以上の試験の補正)
   taud[i,k] <- tau *2*(k-1)/k
   w[i,k] \leftarrow (delta[i,k] - d[t[i,k]] + d[t[i,1]])
                                               #3群以上の臨床試験の補正
   sw[i,k] <- sum(w[i,1:k-1])/(k-1)
                                               #3群以上の臨床試験の累積補正
}
}
totresdev <- sum(resdev[])
                                               # 残差逸脱度の総合計
d[1]<-0
                                               # 参照治療1の治療効果は0
beta[1] <- 0
                                        # 参照治療1の共変量の効果は0
for (k in 2:nt){
                                        # 治療のループ処理
d[k] \sim dnorm(0,.0001)
                                        # 治療効果の漠然事前分布
beta[k] <- B
                                        # 共変量の効果
B \sim dnorm(0..0001)
                                        # 共変量の漠然事前分布
sd \sim dunif(0,5)
                                               # 試験間標準偏差の漠然事前分布
tau <- pow(sd,-2)
                                               # 試験間精度==(1/試験間分散)
                                        # *** PROGRAM ENDS
```

データ構造は事例3と同じだが、ここではセンタリング(中心化)のためにリストデータに共変量の 平均値mxを追加する.

データと連鎖ごとの初期値(変量効果)

```
# Data (BCG vaccine example)
list(ns=13, nt=2, mx=33.46)
                       n[,1] r[,2]
                                                  # ID
t[,1] t[,2]
          na[]
                 r[,1]
                                     n[,2] x[]
    2
                       139
                                     123
                                           44
                                                  # 1
1
                 11
    2
          2
                 29
                       303
                              6
                                     306
                                           55
                                                  #2
1
          2
1
    2
                 11
                       220
                              3
                                     231
                                           42
                                                  #3
(中略)
    2
          2
          2
                       2341
                              5
                                     2498
                                           33
                                                  # 12
    2
                 29
                       17854 27
                                     16913 33
                                                  # 13
1
END
# Initial values
# Initial values for delta can be generated by WinBUGS.
list(d=c(NA, 0), mu=c(0,0,0,0,0,0,0,0,0,0,0,0), sd=1, B=0, sd=1)
#chain 3
list(d=c( NA, 2), mu=c(-3,5,-1,-3,7, -3,-4,-3,-3,0, 5,0,-5), B=5, sd=0.5)
```

共変量の値 0, 13, 及び 50 に対する 2 群の全ての対数オッズ比, オッズ比及び絶対効果を推定するため, ベクトル z を list(z=c(0,13,50), nz=3)として, リストデータに追加できる.

プログラム 4 (b): Binomial likelihood, logit link, 固定効果,連続値の共変量によるメタ回帰分析 (BCG ワクチンの事例)

```
# Binomial likelihood, logit link
# 連続値の共変量を有する固定効果モデル
model{
for(i in 1:ns){
mu[i] \sim dnorm(0,.0001)
for (k in 1:na[i]) {
    r[i,k] \sim dbin(p[i,k],n[i,k])
    logit(p[i,k]) \leftarrow mu[i] + d[t[i,k]] - d[t[i,1]] + (beta[t[i,k]]-beta[t[i,1]]) * (x[i]-mx)
    rhat[i,k] \leftarrow p[i,k] * n[i,k]
    dev[i,k] \leftarrow 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
    + (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
}
resdev[i] <- sum(dev[i,1:na[i]])
}
totresdev <- sum(resdev[])
d[1] < 0
beta[1] <- 0
for (k in 2:nt){
d[k] \sim dnorm(0,.0001)
beta[k] <- B
}
B ~ dnorm(0,.0001)
```

連鎖ごとの初期値(固定効果)

Initial values

#chain 1

list(d=c(NA, 0), mu=c(0,0,0,0,0,0,0,0,0,0,0,0,0), B=0)

#chain 2

#chain 3

list(d=c(NA, 2), mu=c(-3,5,-1,-3,7, -3,-4,-3,-3,0, 5,0,-5), B=5)

固定効果モデルではデータへの当てはまりが、かなりよくなかったので(残差逸脱度の事後平均40と26(<40)データポイントと比較)、変量効果モデルの結果のみを示した(20,000回のBurn-in後、独立な3連鎖から50,000回の繰返しに基づく).表 3-6に共変量「絶対緯度」で調整する場合と調整しない場合の変量効果モデルの結果を示す。共変量で調整するモデルでの治療効果は、共変量の平均値(緯度:33.46°)での効果として解釈することに注意する。異なる緯度での推定された対数オッズ比は、図3-6の実線で表される。

表 3-6 結核 (TB) と診断された患者数に対する共変量を伴わない変量効果モデルに基づく対数オッズ比 (LOR) 及びオッズ比 (OR) (LOR < 0 及び OR < 1 の場合,ワクチンに効果がある),及び交互作用の推定値 (b) の事後平均,標準偏差 (sd) 及び 95%信用区間 (CrI),及び試験間の異質性 (σ) の事後中央値,標準偏差 (sd) 及び 95%信用区間 (CrI),及び活験間の異質性 (σ) の事後中央値,標準偏差 (sd) 及び 95%信用区間 (CrI),及びモデル適合度:残差逸脱度 (resdev) の事後平均,有効パラメータ数 (ρ D) 及び DIC

	·		Mo	odel with Covari	iate†	
	mean/median	sd	95% CrI	mean/median	sd	95% CrI
b	-	-	-	-0.032	0.009	(-0.05, -0.01)
LOR	-0.762	0.220	(-1.21, -0.34)	-0.763	0.126	(-1.04, -0.52)
OR	0.478	0.107	(0.30, 0.71)	0.470	0.059	(0.35, 0.59)
σ	0.649	0.202	(0.39, 1.17)	0.272	0.188	(0.03, 0.75)
resdev*	26.1			30.4		
pD	23.5			21.1		
DIC	49.6			51.5		

^{*26 (}データポイント)と比較, † 共変量「緯度」の平均値(33.46°)での治療効果

20,000 回の Burn-in 後,独立な 3 連鎖から 50,000 回の繰返しに基づく(WinBUGS)

[出典: NICE DSU TSD3 (http://nicedsu.org.uk/technical-support-documents/)]

DICの値を比較すると、共変量で調整したモデルと調整しないモデルで、3又は5未満の差は重要ではないと考えられるため、あまり違いがないように見える―共変量で調整しないモデルでの残差逸脱度の事後平均値はより小さかったが、共変量で調整したモデルでは、変量効果により、より縮小した推定値を許容するため、有効パラメータ数 (pD) がより小さくなる. しかし、共変量で調整するモデルでは、異質性がかなり減少していることがわかる. つまり、共変量で調整しないモデルでは、試験間分散の事後中央値は0.649であるが、共変量で調整するモデルでは0.272となり、交互作用項bの信用区間は零を含んでいない(表 3-6). 共変量で調整するべきかを決定する際には、回帰係数の事後平均値を、事後標準偏差と比較すべきである. DICは変量効果モデルに共変量を含めるかを決定するための信頼できる基準ではない. これは試験間の変動が何であっても、変量効果モデルが同様にデータにうまく当てはまっているためである.

NMA: セルトリズマブ(Certolizumab)の例

NICEの最近の単一の技術評価のアプレイザル(補足:本邦では「総合的評価」と呼ばれる)として、メトトレキセート(Methotrexate; MTX)を含む疾患修飾性抗リウマチ薬(抗リウマチ薬)(Disease-Modifying Anti-Rheumatic Drugs; DMARD)による治療に失敗した患者おける慢性関節リウマチ(Rheumatoid Arthritis; RA)の治療法であるセルトリズマブ(Certolizumab Pegol; CZP)の試験のレビューが実施された[39].

7つの異なる治療法を比較した12つのMTXを対照群とした試験が同定された。すなわち、プラセボ + MTX (治療1) 、CZP + MTX (治療2) 、アダリムマブ+ MTX (治療3) 、エタネルセプト+ MTX (治療4) 、インフリキシマブ+ MTX (治療5) 、 リツキシマブ+ MTX (治療6) 及びトシリズマブ + MTX (治療7) であり、図 3-7に示したネットワークを形成している。

比較が全て同じ治療法に対して行われているこのタイプのネットワークは, しばしば「スター型ネットワーク」と呼ばれる.

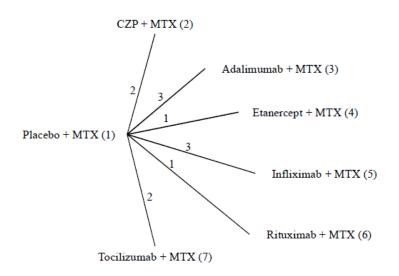


図 3-7 セルトリズマブ (Certolizumab) の例:2 つの治療法を結ぶ線は、これらの治療法を直接比較したことを示している.線上の数字は、2 つの治療法を直接比較した RCT の数である.() 内は NMA での治療番号である.

[出典: NICE DSU TSD3 (http://nicedsu.org.uk/technical-support-documents/)]

表 3-7 に各試験の治療群(i=1,...,12; k=1,2)ごとに、含まれた患者数 n_{ik} 、そのうち 6 ヵ月時点で ARC50 を達成した患者数 (利用可能でない場合は、3 ヵ月時点での ARC50 を使用) r_{ik} 、及び各試験での平均罹患期間 (年) x_i を示した。治療法の相対治療効果に平均罹患期間が影響を与えると考えられ、3.2.1 節のシナリオ B2 に対応している。表 3-7 の未調整のオッズ比 (OR) は、図 3-8 の平均罹患期間に対してプロット (対数尺度) され、2~7 の数字は (参照治療として選択した) プラセボ+ MTX に対する各治療法のオッズ比を示している。Abe 2006 試験での未調整のオッズ比は、表の各セルに 0.5 を加えることにより計算された。

表 3-7 セルトリズマブ(Certolizumab; CZP)例: 12 試験での群 1 及び群 2 での患者の総数 n のうち 6 か月時点で ACR50 に達した患者数 r,及び試験 i の平均罹患期間(年) x_i . 全ての試験でプラセボ又は実薬対照に加えて,MTX が 投与されていた.

			A	Arm 1	A	arm 2	
	Treatment in arm 1	Treatment in arm 2	number achieving ACR50	total number of patients	number achieving ACR50	total number of patients	Mean disease duration
試験名	t_{i1}	t_{i2}	r_{i1}	n_{i1}	r_{i2}	n_{i2}	x_i
RAPID 1	Placebo	CZP	15	199	146	393	6.15
RAPID 2	Placebo	CZP	4	127	80	246	5.85
Kim 2007	Placebo	Adalimumab	9	63	28	65	6.85
DE019	Placebo	Adalimumab	19	200	81	207	10.95
ARMADA	Placebo	Adalimumab	5	62	37	67	11.65
Weinblatt 1999	Placebo	Etanercept	1	30	23	59	13
START	Placebo	Infliximab	33	363	110	360	8.1
ATTEST	Placebo	Infliximab	22	110	61	165	7.85
Abe 2006*	Placebo	Infliximab	0	47	15	49	8.3
Strand 2006	Placebo	Rituximab	5	40	5	40	11.25
CHARISMA*	Placebo	Tocilizumab	14	49	26	50	0.915
OPTION	Placebo	Tocilizumab	22	204	90	205	7.65

^{*3}か月時点のACR50

[出典: NICE DSU TSD3 (http://nicedsu.org.uk/technical-support-documents/technical-support-documents/)]

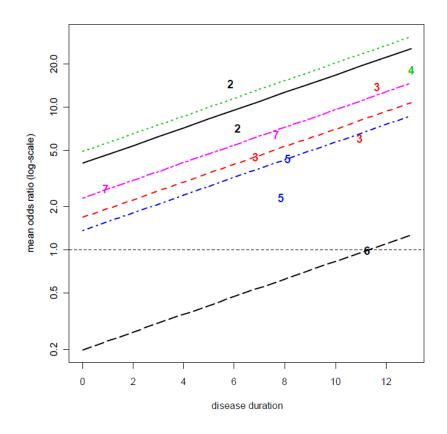


図 3-8 セルトリズマブ (Certolizumab; CZP) の例:6つの実薬対照群のプラセボ+メトトレキセート (MTX) に対する未調整オッズ比 (対数尺度) の平均罹患期間 (年) に対するプロット. プロットされた数字は, プラセボ+MTX と比較した治療を示し, 直線は変量効果メタ回帰モデルに基づいたプラセボ+MTX と比較した以下の治療の相対治療効果を示す. すなわち, (上から下へ) エタネルセプト+MTX (治療 4, 緑の点線), CZP+MTX (治療 2, 黒の実線), トシリズマブ+MTX (治療 7, 紫の長短の破線), アダリムマブ+MTX (治療 3, 赤の破線), インフリキシマブ+MTX (治療 5, 青の一点鎖線), 及びリツキシマブ+MTX (治療 6, 黒の長い破線). 1以上のオッズ比はプロットされた治療が優れていることを示し, 水平線 (薄い破線) の下は治療効果がないことを示す.

[出典: NICE DSU TSD3 (http://nicedsu.org.uk/technical-support-documents/technical-support-documents/)]

すべての治療法に対して共通の交互作用を仮定したモデルを当てはめる. 共通の交互作用項を含む固定効果モデルは, Box 1で述べた. センタリングした (中心化した) 共変量を用いた同等の変量効果モデルを当てはめるには,式(5)を以下のように書き換える.

$$\theta_{ik} = \text{logit}(p_{ik}) = \mu_i + \left(\delta_{i,1k} + (\beta_{1t_{ik}} - \beta_{1t_{ij}})(x_i - \bar{x})\right) I_{\{k \neq 1\}}$$
 (7)

ここで,

$$\bar{x} = 8.21, \beta_{11} = 0, \beta_{1k} = b(k = 2, ..., 7)$$

$$\delta_{i,1k} \sim N(d_{1t_{ik}}, \sigma^2)$$

である.

本モデルは多くの方法で表現でき、コンピュータでの実装のためにコード化することができる.この 定式化では、治療法に特有の交互作用は保持するが、それらはすべてbに等しく設定する.この定式化に より、実薬群同士の比較で、交互作用項が相殺される.この定式化のコードをセルトリズマブの事例で 与えた.

最後に、2群比較のメタアナリシスは、NMAの特別なケースであるため、2つの治療法のみの場合、式(7)のモデルは式(6)のモデルに簡略化されることに注意する.

基本パラメータ d_{1k} 及びbには、無情報の正規事前分布が与えられる。試験間の標準偏差の事前分布,及び対応するWinBUGSコードの詳細を以下に示す。連続値の共変量及び無情報事前分布を持つメタ回帰分析による変量効果モデルのWinBUGSコードはプログラム4(a)に、固定効果のコードはプログラム4(b)に与えた。モニターすべき関連するパラメータはプログラム3と同じである。データ構造は事例4と同じだが、今回は2つ以上の治療法を比較する。

データと連鎖ごとの初期値(変量効果)

`	is=12, n	it=7, mx=	8.21)						
t[,1]	t[,2]	na[]	n[,1]	n[,2]	r[,1]	r[,2]	x[]	# ID Study name	
1	3	2	63	65	9	28	6.85	# 1 Kim 2007 (37)	
1	3	2	200	207	19	81	10.95	# 2 DE019 Trial (36)	
1	3	2	62	67	5	37	11.65	# 3 ARMADA Trial (34)	
(中	略)								
1	7	2	49	50	14	26	0.92	# 12 CHARISMA Study (64)	
1	7	2	204	205	22	90	7.65	# 13 OPTION Trial (67)	
END)								
# Ini	tial valu	es for RE	model						
	in 1								
#cha					0 0 0 0		0.0) cd_1	D 0/	
#cha		, 0,0,0,0,0	0,0), mu=	c(0, 0, 0,	0, 0, 0, 0	J, U, U, U,	0,0), Su= i	, D=U)	
#cha	l=c(NA,	, 0,0,0,0,0	0,0), mu=	c(0, 0, 0,	0, 0, 0, 0), 0, 0, 0,	0,0), Su=1	, b=0)	

連鎖ごとの初期値(固定効果)

 σ の事前分布として Uniform (0,5) を用いた変量効果モデルでは,異質性パラメータは特定できない.これはゼロセル(補足:発現例数又は非発現例数のいずれかが 0)の試験が 1 つあり,各比較の繰返し数が多くないためである.試験間の変動を推定可能となるための情報が不足しており, σ の事前分布が情報のある事前分布でない場合,本試験の相対治療効果は無限大に向かう.そのため,図 3-9 の実線で表されている,安定した計算を保証するために,情報のある事前分布として半正規分布を用いた.

$\sigma \sim \text{Half-Normal}(0,0.32^2)$

この事前分布は、各比較に対して、試験特有のオッズ比(OR)の 95%が、OR の中央値から 2 倍以内になるように、事前に、選択された。この事前分布のもとで、 σ の事後平均値は 0.26 となる。この事前分布をメタ回帰分析による変量効果モデルに適用するには、プログラム 4 (a) で「試験間の SD について漠然事前分布」と注釈が付けられているコード行を、以下の 2 行で置き換える必要がある。

sd ~ dnorm(0,prec)I(0,) # 試験間標準偏差の事前分布 prec <- pow(0.32,-2)

本事前分布は何も考えずに使用されるべきではない。上記のプログラムコードで prec 値を変更することで、広い範囲又は狭い範囲の値を許容できる情報のある事前分布が使用できる。本事例では、 σ の事後分布は、図 3-9 の点線で示されており、 σ の妥当な値の範囲は大きく変わらないことを示すが、 σ が 0 に近い値での確率が減少していることが判る。

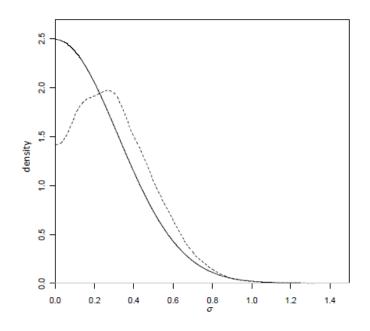


図 3-9 セルトリズマブ (Certolizumab) の事例:情報のある事前分布 Half-Normal (0,0.32²) によるメタ回帰分析. 事前分布の確率密度関数は実線で、事後確率密度関数は点線で表される.

[出典: NICE DSU TSD3 (http://nicedsu.org.uk/technical-support-documents/ (http://nicedsu.org.uk/technical-support-documents/ (http://nicedsu.org.uk/technical-support-documents/ (http://nicedsu.org.uk/technical-support-documents/ (http://n

解析では中心化した共変量の値が用いられるため、得られた治療効果は、共変量の平均値(事例では 8.21 年)において推定された対数オッズ比であり、共変量の値 z での推定値を算出するために、中心化をやめて、変換することで $d_{1k}-b(\bar{x}-z),k=2,...,7$ として求められる.

表 3-8に固定効果及び変量効果のNMA,及び罹患期間を共変量とする交互作用モデルの結果を示す (40,000回のBurn-in後,独立した3つの連鎖から100,000回の繰返しに基づく).様々な罹患期間に対する推定オッズ比は、図 3-8の平行線で表される.

表 3-8 セルトリズマブ(Certolizumab)の例: ACR50 に達した患者数に対する共変量「罹患期間」がある場合とない場合での固定効果及び変量効果モデルに基づいた,交互作用の推定値(b)及び治療 1 に対する治療 Y($2\sim7$)の対数 オッズ比 d_{1Y} の事後平均,標準偏差(sd)及び 95%信用区間(CrI),及び試験間の異質性(σ)の事後中央値,標準偏差(sd)及び 95%信用区間(CrI),並びにモデル適合度:残差逸脱度の事後平均(resdev),有効パラメータ数(pD)及び DIC.

	No covariate						Covariate 'disease duration'					
	Fixed effects			Random effects†			Fixed effects			Random effects†		
	mean/	sd	CrI	mean/	sd	CrI	mean/	sd	CrI	mean/	sd	CrI
	median			median			median			median		
b	-	-	-	-	-	-	0.14	0.06	(0.01, 0.26)	0.14	0.09	(-0.03, 0.32)
d_{12}	2.21	0.25	(1.73, 2.72)	2.27	0.39	(1.53, 3.10)	2.50	0.29	(1.96, 3.08)	2.57	0.42	(1.79, 3.44)
d_{13}	1.93	0.22	(1.52, 2.37)	1.97	0.33	(1.33, 2.64)	1.66	0.25	(1.19, 2.16)	1.71	0.34	(1.04, 2.41)
d_{14}	3.47	1.34	(1.45, 6.74)	3.46	1.41	(1.26, 6.63)	2.82	1.34	(0.71, 5.96)	2.77	1.42	(0.42, 6.01)
d_{15}	1.38	0.17	(1.06, 1.72)	1.48	0.33	(0.90, 2.21)	1.40	0.17	(1.08, 1.74)	1.48	0.30	(0.95, 2.15)
d_{16}	0.00	0.71	(-1.40, 1.39)	0.01	0.82	(-1.61, 1.63)	-0.42	0.73	(-1.86, 1.04)	-0.44	0.84	(-2.08, 1.21)
d_{17}	1.65	0.22	(1.22, 2.10)	1.56	0.38	(0.77, 2.28)	1.98	0.28	(1.45, 2.53)	2.00	0.45	(1.12, 2.93)
σ	-	-	-	0.34	0.20	(0.03, 0.77)	-	-	-	0.28	0.19	(0.02, 0.73)
resdev*	37.6			30.9			33.8			30.2		
pD	18.0			21.2			19.0			21.3		
DIC	55.6			52.1			52.8			51.4		

*24 (データポイント) と比較、 $\dagger \sigma$ に対し情報のある事前分布(半正規分布)を利用 40,000 回の Burn-in 後、独立した 3 つの連鎖から 100,000 回の繰返しに基づく(WinBUGS)

[出典: NICE DSU TSD3 (http://nicedsu.org.uk/technical-support-documents/)]

表 3-8の4つのモデルに対するDIC及び残差逸脱度の事後平均値は、単一のモデルを決定的に支持するものではなかった。固定効果モデルのみを比較すると、共変量との交互作用項b,その信用区間は零を含まない、を含めることで当てはまりが改善することがわかる。共変量を含めた変量効果モデルは異質性を減少させるが(共変量を含めない変量効果モデルでの試験間分散の事後中央値は0.34であるが、共変量を含めた変量効果モデルの事後中央値は0.28であった)、交互作用パラメータbの信用区間は零を含んでいた。メタ回帰モデルは完全に合理的であるが、エビデンスにより強く支持されなかった。それにもかかわらず、より短い罹患期間で、より小さな治療効果が認められることは、より大きな(複数の)試験で報告されており[37]、決定モデルの意味合いを考慮する必要がある。本問題は、生物(学的)製剤の使用を、罹患期間が特定の関値を超えている患者に限定すべきか否かである。それは不合理な考え方ではないが、図 3-8のメタ回帰のみに基づいて閾値を決定することは難しいだろう。治療法3及び7(アダリムマブ及びトシリズマブ)、これらは1つ以上の罹患期間で検討された治療法であり、各期間で異なる治療効果を示しているように見え、回帰の傾きの決定に大きく影響を与えている。したがって、線形の関係性は非常に疑わしく、また治療法6(リツキシマブ)での負の治療効果の予測は妥当ではない、用いたメタ回帰モデルが妥当ではなく、異質性の原因に関して、更なる調査が行われるべきことを示唆している(3.7.4節参照)。

3.7.4 ベースラインリスクによるメタ回帰

ベースラインリスクに関するメタ回帰モデルは、式(7)と同じであるが、共変量は各試験での対照群での試験固有のベースライン(リスク)となる($x_i = \mu_i$).

再掲載:
$$\theta_{ik} = \text{logit}(p_{ik}) = \mu_i + (\delta_{i,1k} + (\beta_{1t_{ik}} - \beta_{1t_{i1}})(x_i - \bar{x}))I_{\{k \neq 1\}}$$
 (7)

このベイズ流の定式化での重要な特性は、(モデルで推定される)「真の」ベースライン(リスク)を共変量とみなし、各 μ_i の不確実性を自動的に考慮に入れる点である[40][41]. 観察されたベースラインリスクに対して後退する(回帰する)ナイーブなアプローチは、治療効果とベースラインリスクとの相関関係を考慮しないため、平均への回帰現象を引き起こしてしまう。

共変量 μ_i は、線形予測子(例えば、logit、log又はidentity(恒等)尺度)と同じ尺度であることに注意することが重要であり、センタリングのための共変量の平均値もこの尺度である必要がある。例えば、logitリンク関数を利用する場合、各試験のベースライン群(k=1)の平均対数オッズを μ_i から引くことにより、共変量をセンタリングする必要がある。NMAでは各試験でのベースライン群の治療が常に治療1(基準治療)であるわけではない(治療1は含まれず、別の治療がベースライン群になっている可能性がある)ことに注意すること。しかし、式(7)のモデル(Box 1でのモデル3)では、治療1と比較する全ての治療群で同じ交互作用が仮定されているため、その交互作用項は、他の全ての比較(治療1以外の治療群の対比較)で相殺されるため、治療1を含まない試験に対しては、ベースラインリスクの調整は行われない。一方で、Box 1のモデル1、2を当てはめる場合、調整されるリスクが、全ての試験で比較されなかった(かもしれない)基準治療(治療1)の推定リスクを参照するように注意する必要がある。

ベースラインリスクに関するネットワークメタ回帰:セルトリズマブ(Certolizumab)の例

図 3-10は、セルトリズマブ(Certolizumab)の例について、表 3-7から得られた未調整のオッズ比を(対数尺度で)ACR50のベースラインオッズに対してプロットを示す。No.2~7は、プラセボとMTX(基準治療として選択)に対するその治療のオッズ比を表す。プロットの目的のために、Abe 2006 試験の未調整のオッズ比は、各セルに0.5を加えて計算され、ベースライン対数オッズは0.01と推定された。図 3-10は、治療効果とベースラインリスク(対数尺度)との間に強い線形関係を示唆しているに見える。3.7.3節で論じたように、式(7)のモデルでは、図 3-10の点に平行な回帰直線が当てはまると仮定しており、回帰直線間の差がベースラインリスクを調整した真の平均治療効果を表している。

共通の交互作用項を有する固定効果モデルと変量効果モデルの両方を当てはめられた。基本パラメータ d_{1k} とbには,無情報事前分布として,正規分布の $N(0,100^2)$ と(一様分布) σ ~Uniform (0,5) が与えられる。ベースラインリスクに対するメタ回帰のWinBUGSコードを以下に示す。ベースラインリスクの調整を伴うメタ回帰分析における変量効果及び固定効果モデルのWinBUGSプログラムコードは,それぞれプログラム4(a)及び4(b)と同様である。しかし,線形予測子の定義により,x[i]がmu[i]により置き換えられる。正規分布の事前分布のバラツキは,数値計算での誤差を避けるために減少させる必要がある(この対処は,事後分布の結果に最小限の影響しか与えない)。

プログラム 6 (a): Binomial likelihood, logit link,変量効果,ベースラインリスクで調整するメタ回帰分析(セルトリズマブ(Certolizumab)の事例)

```
# Binomial likelihood, logit link
# Random effects model for multi-arm trials
                                                  # *** PROGRAM STARTS
model{
for(i in 1:ns){
                                          # 試験のループ処理
                                          #3群以上の臨床試験の補正(対照群は0)
w[i,1] < 0
                                          # 対照群の治療効果は0
delta[i,1] <- 0
                                          # ベースラインの漠然事前分布
mu[i] \sim dnorm(0, \underline{.001})
for (k in 1:na[i]) {
                                                  # 治療群のループ処理
   r[i,k] \sim dbin(p[i,k],n[i,k])
                                                  # 二項尤度
   # 線形予測子に対するモデル, 1つ目の群に対する共変量の効果
   logit(p[i,k]) \leftarrow mu[i] + delta[i,k] + (beta[t[i,k]]-beta[t[i,1]]) * (mu[i]-mx)
   rhat[i,k] \leftarrow p[i,k] * n[i,k]
                                                  # 分子の期待値
   dev[i,k] \leftarrow 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
                                                  # 逸脱度
   + (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
                                                  # 試験ごとの残差逸脱度の合計
resdev[i] <- sum(dev[i,1:na[i]])
for (k in 2:na[i]) {
                                                  # 治療群のループ処理
                                                  # 試験特有の対数オッズ比(LOR)の分布
   delta[i,k] ~ dnorm(md[i,k],taud[i,k])
   md[i,k] \leftarrow d[t[i,k]] - d[t[i,1]] + sw[i,k]
                                                  #LOR事後分布の平均(3群以上の試験の補正)
   taud[i,k] <- tau *2*(k-1)/k
                                                  #LOR事後分布の精度 (3群以上の試験の補正)
                                                  #3群以上の臨床試験の補正
   w[i,k] \leftarrow (delta[i,k] - d[t[i,k]] + d[t[i,1]])
   sw[i,k] <- sum(w[i,1:k-1])/(k-1)
                                                  #3群以上の臨床試験の累積補正
}
}
                                                  # 残差逸脱度の総合計
totresdev <- sum(resdev[])
d[1]<-0
                                                  # 参照治療1の治療効果は0
beta[1] <- 0
                                          # 参照治療1の共変量の効果は0
for (k in 2:nt){
                                          # 治療のループ処理
d[k] \sim dnorm(0,.0001)
                                          # 治療効果の漠然事前分布
beta[k] <- B
                                          # 共通の共変量効果
}
B ~ dnorm(0,.0001)
                                          # 共変量効果の漠然事前分布
sd \sim dunif(0,5)
                                                  # 試験間標準偏差の漠然事前分布
tau <- pow(sd,-2)
                                                  # 試験間精度=(1/試験間分散)
                                          # *** PROGRAM ENDS
}
```

データ

```
# Data (Certolizumab, baseline risk)
list(ns=12, nt=7, mx=-2.421)
          na[] n[,1]
                                                  # ID Study name
t[,1] t[,2]
                           n[,2]
                                   r[,1]
                                           r[,2]
            2
1
     3
                    63
                           65
                                   9
                                           28
                                                  # 1 Kim 2007 (37)
     3
                    200
                           207
                                   19
                                           81
                                                  # 2 DE019 Trial (36)
1
     3
            2
                    62
                           67
                                          37
                                                  #3 ARMADA Trial (34)
                                   5
 (中略)
     7
            2
                    49
                           50
                                   14
                                           26
                                                  # 12 CHARISMA Study (64)
                    204
                           205
                                                  # 13 OPTION Trial (67)
1
END
```

連鎖ごとの初期値(変量効果)

プログラム 6 (b): Binomial likelihood, logit link, 固定効果, ベースラインリスクで調整するメタ回 帰分析 (セルトリズマブ (Certolizumab) の事例)

```
# Binomial likelihood, logit link
# Fixed effects model with one covariate (independent covariate effects)
model{
for(i in 1:ns){
mu[i] \sim dnorm(0, .001)
for (k in 1:na[i]) {
    r[i,k] \sim dbin(p[i,k],n[i,k])
    logit(p[i,k]) \leftarrow mu[i] + d[t[i,k]] - d[t[i,1]] + (beta[t[i,k]]-beta[t[i,1]]) * (mu[i]-mx)
    rhat[i,k] \leftarrow p[i,k] * n[i,k]
    dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
    + (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
}
resdev[i] <- sum(dev[i,1:na[i]])
}
totresdev <- sum(resdev[])
d[1] < 0
beta[1] <- 0
for (k in 2:nt){
d[k] \sim dnorm(0,.0001)
beta[k] <- B
B \sim dnorm(0..0001)
}
```

連鎖ごとの初期値(固定効果)

推定された μ_i のそれぞれから共変量の平均値(つまり、治療 1 での観察対数オッズの平均値、 $\bar{x}=-2.421$)を差し引くことにより、センタリングを行っている(本試験はいずれもベースライン群が治療 1 であったため、すべての試験において μ_i から治療 1 での対数オッズの平均値を引いている).得られた治療効果 d_{ik} は、共変量の平均値($z=\bar{x}$)での推定対数オッズ比であるが、 $d_{ik}-b(\bar{x}-z)$ 、

k=2,...,7 とすると、任意のベースラインリスクzでの治療効果の推定値に変換できる.表 3-9 は、共変量としてのベースラインリスクを有し、固定効果及び変量効果モデルでの交互作用の検討結果を示す(結果は 60,000 回の Burn-in 後の 3 つの独立した連鎖からの 100,000 回の反復に基づく).共変量調整を伴うモデルでの治療効果は、ACR50 のベースラインロジット確率-2.421 を逆ロジット関数(TSD2[8])を用いて、ACR50 のベースライン確率 0.082 に変換することができ、(そのベースライン確率を有する)患者の効果として解釈される.

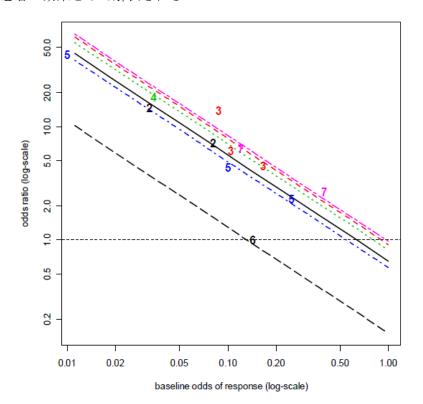


図 3-10 セルトリズマブ (Certolizumab; CZP) の例:ベースラインオッズ (対数尺度) に対して、プラセボ+メトトレキセート (MTX) に対する6つの実薬での治療の未調整オッズ比のプロット. プロットされた数字は、プラセボ+MTX と比較される治療を表し、直線は変量効果メタ回帰モデルに基づくプラセボ+MTX と比較して、以下の治療の相対治療効果を表している。すなわち、(上から下へ)トシリズマブ+MTX (7、紫の短長破線)、アダリムマブ+MTX (3、赤の破線)、エタネルセプト+MTX (4、緑の点線)、CZP+MTX (2、黒い実線)、インフリキシマブ+MTX (5、濃紺の点と破線)、リツキシマブ+MTX (6、黒い長破線).1以上のオッズ比はプロットされた治療が有効であり、水平線(破線)以下は治療効果がないことを表す。

[出典: NICE DSU TSD3 (http://nicedsu.org.uk/technical-support-documents/technical-support-documents/)]

表 3-9 セルトリズマブ (Certolizumab) の例:交互作用の推定値 (b) 及び治療 X に対する治療 Y の対数オッズ比 d_{XY} の事後平均値,標準偏差 (sd) 及び 95%信用区間 (CrI). 試験間の異質性 (σ) の事後中央値,sd 及び 95%信用区間 (CrI). 「ベースラインリスク」の共変量を含める/含めない固定効果モデル及び変量効果モデルとして検討. モデルの 適合度:残差逸脱度の事後平均 (resdev),有効パラメータの数 (pD) 及び DIC

		Fixed Effects			Random Effect	s
	mean	sd	CrI	mean	sd	CrI
b	-0.93	0.09	(-1.03,-0.69)	-0.95	0.10	(-1.10,-0.70)
d_{12}	1.85	0.10	(1.67, 2.06)	1.83	0.24	(1.35,2.29)
d_{13}	2.13	0.11	(1.90,2.35)	2.18	0.22	(1.79,2.63)
d_{14}	2.08	0.34	(1.47,2.80)	2.04	0.46	(1.19,2.94)
d_{15}	1.68	0.10	(1.49, 1.86)	1.71	0.22	(1.30,2.16)
d_{16}	0.36	0.50	(-0.72,1.27)	0.37	0.59	(-0.86,1.45)
d_{17}	2.20	0.14	(1.93,2.46)	2.25	0.27	(1.75,2.79)
σ	-	-	-	0.19	0.19	(0.01, 0.70)
resdev*	27.3			24.2		
pD	15.0 [†]			19.4		
DIC	42.3 [†]			43.6		

^{*24 (}データポイント) と比較

†公開されているWinBUGSのプログラムに保存されている結果から、出典の表の転記ミスと思われる数値を訂正している。ただしTSD3の著者らには確認はしていない。

60,000回のBurn-in後の3つの独立した連鎖からの100,000回の反復に基づく(WinBUGS)

[出典: NICE DSU TSD3 (一部改変) (http://nicedsu.org.uk/technical-support-documents/)]

共変量を伴う固定効果モデルと変量効果モデルとも、交互作用項の信用区間が0からかなり遠いため、ベースラインリスクと治療効果との強い交互作用効果を示唆している。ベースラインリスクと治療効果との交互作用を伴う変量効果モデルでは、異なる治療法に対する推定オッズ比は、図 3-10の異なる平行線により表される。残差逸脱度の事後平均値は、変量効果モデルがわずかに優れているように見えるが、DICの差はほとんどなく(前述したように、DICの2~3の差は小さいと見なされる)、よりシンプルな固定効果モデルを用いることが適切と考えられる。

注目すべき重要な点は、回帰式での共通の交互作用項bを仮定することで、1つの試験しかない比較でも交互作用パラメータを推定できるということである。また、ある試験で測定された範囲外のベースラインリスク値での治療効果の推定も可能となる。例えば、リツキシマブ+MTX(治療6)とプラセボ+MTXとの比較試験は1つのみである。モデルの仮定は、他のものと平行となる線が、この点を通って描かれることを示唆している(図 3-10)。また、この分析はMTXにリツキシマブを加えることは、他の治療よりも患者にはあまり有益ではないことを示唆しており、ベースラインリスクが0.15以上であれば、恐らくもっともらしくないが、有害である可能性があると予測している。

恐らく治療6を除いた(当てはまりが良いかは、1試験のみのため判断ができない)、全ての治療に対して1つの交互作用項を用いた図 3-10の顕著な当てはまりの良さは、意思決定及び実際のエビデンス統合において、幾つかの示唆がある。初めに、CEAモデルに組み込む必要がある有効性とベースラインリスクとの関係を明確に示唆している。次に、図 3-10は、2群比較のメタアナリシスでの深刻な異質性、及

びNMAでの潜在的な不一致を導入する共変量があり、もしその共変量を制御できない場合、それらに起因した治療効果の大きさ(effect size)が、どのように変動するかを示している。ベースラインリスクが考慮されると、試験間(治療内)と抗TNF-α薬物間の差異の両方が最小限であることは明らかである.

3.7.5 IPD を利用したメタ回帰

個々の患者データ (IPD) のメタアナリシスはゴールドスタンダード[42]と述べられており、解析方法 の標準化の可能性を含み、要約されたデータのエビデンス統合よりも一定の利点がある[43]. さらに、患者レベルの共変量に関心があり、個々の患者特性を個々の患者のアウトカムに回帰するためにIPDを 利用することは、3.7.1節~3.7.4節で考察された集計アウトカム及び (集計) 共変量データの利用と比較して、より強力で信頼性の高い分析をもたらす[44][45]. そのような解析は、通常、集計データに基づく分析よりもはるかに強力であるだけでなく、潜在的なエコロジカルバイアスを回避可能となる. 連続値の共変量及び連続値のアウトカムを扱う場合、IPDメタ回帰分析が本質的である (不可欠である).

以下では、IPDが全ての試験で利用可能である状況と、一部の試験でのみ利用可能な状況を区別している.

患者レベルの共変量を利用する方法(全ての試験で IPD が利用可能)

IPDを用いたメタアナリシスでは、歴史的に、広い意味で1ステップアプローチと2ステップアプローチからなる2つのアプローチが検討されている[46]. 便宜上、1ステップアプローチは後述する.

2ステップアプローチでは、ステップ1として、各試験から得られる興味のある効果の大きさを、不確実性の指標(例えば、標準誤差)と共に推定し、ステップ2として、その要約データを用いて、標準的方法でメタアナリシスを行う。この場合、共変量と治療との交互作用を検討することで治療効果の異質性を検討する[47]. 状況によっては、分析者が全てのIPDにアクセスできない(一部のIPDにアクセスできない)場合でも、このようなIPD分析を実行することは可能かもしれない。すなわち、データの所有者は、IPDデータセット全体を提供したくない場合、そのような交互作用の効果を計算して提供するかもしれない。しかし、複数の共変量を同時に考慮する必要がある場合、そのようなアプローチは煩雑であるか、実行不能になる(補足:まず調整すべき因子が事前にすべて明らかになっているという前提であるが(そのようなことはほとんどないかもしれない)、データ提供者がデータを提供したくはないが、複雑な解析には合意した場合、すべての共変量/因子を調整した上で、1つのサブグループ変数によるサブグループ解析結果により治療効果の異質性が検出できる。しかしながら、共変量/因子同士の関係性(交互作用)を詳細に検討する等は検討できなくなる)。

2ステップアプローチは交互作用の存在に対する推論には有用だが、パラメータ間の相関をモデルを 通して反映できるよう、主効果と交互作用を同時に推定する必要がある意思決定には有益ではない.

1ステップアプローチでは、全てのIPDが単一のデータセットにまとめられ、データ内での試験内比較を維持しながら、すべてのデータを同時に分析するものである.2群比較のメタアナリシスでのIPD変量効果モデルは、連続値[48][49]、2値[50]、生存時間データ[51]及び順序データ[52]のアウトカムに対して

開発されており、全ての患者レベルの共変量を含めることができる. 1ステップアプローチの適用例として、ほとんどのモデルは従来の対比較で利用されてきたが、NMAにそれらを拡張することが可能である[46][53]. また、患者レベルの共変量を評価するIPDの潜在的利益を決定するための単純な基準を考慮しているものもある[47].

したがって、要約データのみが利用可能である場合、共変量と治療効果との交互作用は、試験間のバラツキの情報を利用可能であり(メタ回帰)、またもしIPDが利用可能である場合は、試験内のバラツキ情報を利用して推定可能である。しかし、IPDを利用する場合の優れた点は、試験間及び試験内の(回帰)係数の両方を推定可能であることである[48]。これは、2種類の共変量をモデルに含めることで達成できる。すなわち、1つ目の共変量は、その試験での共変量の平均値(すなわち、1つの試験で各患者が有する値、つまり同じ要約統計量を用いるが、メタ回帰分析を実施する場合に使用する値)であり、2つ目の共変量は個々の患者応答からその試験の平均値を差し引いたものである。具体的で詳細なモデル化は別のところで利用可能である[54]。これは連続値の共変量に最も自然に適用されるが、2値の共変量にも適用できることに注意する(例えば、2値の共変量が性別である場合、試験レベルの共変量は女性の割合となる)。

これらの二重の効果(試験内及び試験間)のモデルを利用するには、多くの方法がある。最も魅力的な選択肢は、エコロジカルバイアス/要約バイアス(aggregation bias)や他の試験間の潜在的な交絡要素がないため、試験内の変動(バラツキ)からのみ得られた推定値を利用することである。潜在的に、試験内及び試験間の共変量で同じパラメータを有することで、試験間の変動(バラツキ)に情報を反映することにより、検出力を向上させることができる。もちろん、これも同じバイアスを潜在的に誘発し得る可能性がある。2つの推定値間(試験内と試験間)の差異の統計的検定を実施することができ、試験内及び試験間の共変量で、同じ交互作用があるかの決定は、この検定に基づき得られることが示唆されている[54]。しかし、本検定は多くの状況で検出力が低い可能性があり、使用を推奨するには、このアプローチの更なる検討が必要である。

IPD と要約データの組み合わせの利用(一部の試験で IPD が利用可能)

IPDが、全てではないが多数の関連する試験から利用可能である状況が存在する可能性がある。そのような場合、2群比較のメタアナリシスでは、異質性を探索するための利用可能な3つの潜在的な選択肢がある。1つ目は、IPDが利用できない全ての試験を除外することであり。これは解析を単純にし、(前節で説明した方法を利用して)試験内の比較のみに基づいて行うことができるが、関連する全ての試験を含めることができない欠点がある。さらに、IPDを提供しない理由が治療効果に関連する場合、解析は潜在的に偏っている可能性がある。2つ目は、要約データのメタ回帰を実行することである。これは潜在的に全ての臨床試験が含まれる可能性があるが、一部のIPDを有することの利点を諦めることになる。最後に、利用可能であればIPDを組み入れ、利用可能でない場合は要約データを組み入れることを可能にしたモデルが開発されており[55]、そのモデルを利用することである。このアプローチは、全てのデータを、全ての試験から入手可能な最も詳細なレベルで含めることを可能にするが、IPDのみの分析では、試験間のバラツキを効果の推定に含めるかどうかを決定しなければならない。ここでもまた、試験間及び試験内のバラツキを用いて、効果の差の検定を構築でき、どのアプローチをとるかを決定するために

利用することができる(ただし、ここでもまた、多くの状況で、恐らく検出力が低いことが問題となる可能性に注意する). IPD及び要約データの組み込みを可能にするモデルが、2値アウトカム[55][56]及び連続値アウトカム[57]に対して提案されている.

これまでのNMAで、IPDと要約データを同時に利用する研究はほとんどなかった。全ての試験において幾つかの群間比較に関してはIPDが利用可能であるが、他の比較ではIPDが利用可能な試験が1つもないことは、非常に考えられることである。これは、企業が自社製品の臨床試験データに完全にアクセスできるが、競合他社製品では要約データのみにアクセスできるような状況であり、製薬業界で行われている個別技術評価(STA: Single Technology Appraisal)では特に当てはまる可能性がある。3.7.2節で説明したように、プラセボ又は標準治療(usual care)との交互作用の影響が同じであるか、交換可能であるか、又は治療によって異なるかを決定しなければならない。TSDでは同じクラスのすべての治療に対し、1つの交互作用パラメータを使用することを示唆しているが、すべての可能性に対してモデルを構築することはできる。また、試験内及び試験間の2つの共変量に対するモデルへの拡張が可能であり、これを最初に検討した研究がある[58]。

幾つかの異なる治療に対して、IPDが利用可能である場合は、患者レベルの交互作用が治療を通して同じかを、より完全に検討することを可能にする.

なお,2016年12月にTSD18 (Methods for population-adjusted indirect comparisons in submissions to NICE: http://nicedsu.org.uk/wp-content/uploads/2017/05/Population-adjustment-TSD-FINAL.pdf) がリリースされており、IPDデータを利用した間接比較が取り扱われている。2020年度のDS部会TF6にてTSD18 を紹介する報告書を成果物としてリリースすることを検討中である。

3.8 バイアス(偏り)とその調整

本節では、内的バイアス(内的妥当性を脅かすバイアス)と外的バイアス(外的妥当性を脅かすバイアス)の両方に対して、バイアス補正方法を検討する。前述した「バイアス補正」とメタ回帰モデルの違いはわずかであるが、重要である。メタ回帰では、意思決定問題の正式な範囲内でさえ、相対治療効果に明確な違いがあることを認める。バイアス補正では、意思決定のための対象集団を念頭に置いているが、利用可能なエビデンス又は少なくとも幾つかのエビデンスは、恐らく試験が内的バイアスを有するため、恐らく異なる集団又は異なる設定、又はその両方に関係しているため、対象となるパラメータの「偏った」又は潜在的に偏った推定値を与える。Box 2は、バイアスを補正する4つのアプローチをまとめたもので、詳細は以下で述べる。

通常、回帰は共変量の違いを調整する1つの方法と見なされるが、共変量が外的妥当性に影響を及ぼし、共変量がバイアス補正の問題として扱われているため、それを「バイアス補正」の方法と呼ぶこととする.

Box 2 バイアスを補正する異なるアプローチ

- メタ回帰による共変量の調整 (3.7節):具体的な対象集団及び治療法の決定が必要であるが、多くのエビデンスには、他の集団又は他の(類似の)治療が含まれる.このアプローチは、2群比較のメタアナリシス、間接比較及び無作為化比較試験のNMAに適しており、より大きなデータセットでより効果的である.
- メタ疫学研究データに基づくバイアスの補正[60]:エビデンスには潜在的バイアスを示す幾つかの試験が含まれており、そのバイアスの事前分布を外部のメタ疫学研究の情報から推定できる.このアプローチは、2群比較でのメタアナリシス、間接比較、NMA及び「品質」が混合した無作為化比較試験に適しているが、臨床試験と観察研究データが混在したメタアナリシスにも拡張できる.このアプローチは、各比較に対して1つの試験のみの場合を含む小さなデータセットには適しているが、利用されるメタ疫学研究からのデータの適切性に依存する.
- 試験ネットワークでのバイアスの推定と調整 [61][62][63]:バイアスの程度は、既存の試験エビデンスから内的に推定可能である.このアプローチは、間接比較又は「品質」が混合した無作為化比較試験のNMAに適しているが、臨床試験と観察研究データの混合に拡張することができる.より大きなデータセットではより効果的である.
- 専門家又はデータに基づいたバイアス分布の導出[59]:上記のいずれの状況にも適用でき,2群比較のメタアナリシス,間接比較,無作為化比較試験又は観察研究のNMAに適している[64].このアプローチは,1つの試験を含む小さなデータセットには適しているが,非常に時間がかかる場合がある.

3.8.1 メタ回帰による共変量の調整

この方法は前出のメタ回帰モデルを利用する.このアプローチは、対象集団又は治療の治療効果に関するエビデンス(臨床試験の数又は過去のメタアナリシスの存在)が限られているが,他の関連する集団又は関連する治療の情報が存在する場合のオプションである.2つのグループで治療効果が系統的に異なると感じた場合、メタ回帰分析は、追加の関連する一連の臨床試験から「力を借りる」方法となる.例えば、RA(リウマチ)治療薬の生物学的製剤の場合、非生物学的製剤のDMARD [疾患修飾性抗リウマチ薬]による治療には失敗したが、MTXを服用することはできない患者への治療の決定が必要と仮定する.理想的には、生物学的製剤とプラセボを含む治験が必要である.この患者群で利用可能であるデータが少ない場合、生物学的製剤+MTXとプラセボ+MTXを比較したより多くのデータを利用することが可能かもしれない.MTXの存在下でのプラセボに対する生物学的製剤の共通の追加効果を調整する交互作用モデルを用いて、この追加のデータから力を借りることができる.この場合、調整は生物学的製剤間の比較に対してではなく、生物学的製剤とプラセボの比較にのみ関係することに注意する(Box 1参照).

また、MTXの服用の有無にかかわらず、交互作用のない、すなわちプラセボに対する生物学的製剤の効果は同じであったと仮定する選択肢もある。この場合、データ全体を、交互作用項を導入せずに、プラセボに対する生物学的製剤の治療効果、及び生物製剤同士の治療効果を推定するために利用できる。

3.8.2 メタ疫学研究データに基づくバイアスの補正

Schulzらは、「質の高い」無作為化比較試験の結果を、潜在的に低品質の指標、すなわち盲検化の欠如 又は二重盲検化の欠如した試験の結果と比較した。彼らのデータセットには30以上のメタアナリシスが 含まれており、それぞれに「高」と「低」の両方の品質の試験が存在した。彼らの結果は、新しい治療 法に有利な相対治療効果は、品質の低い試験で(平均して)より高いことを示唆していた。その影響は 大きく、オッズ比は平均で約1.6倍高く新しい治療に有利であった。

様々な品質のエビデンスが混在している場合、研究者には2つのオプションがあった:質の高い試験に限定するか、あるいは高品質と低品質の全ての試験を1つの分析に含めるかである。両方の選択肢には欠点がある:最初のものは、エビデンスのかなりの部分を無視しており、2つ目は、治療効果の偏った推定値をもたらす危険性がある。

Weltonら[60]は、全てのデータを使用するアプローチを示唆しているが、質の低い試験からのエビデンスを同時に調整し、重みを低くしている.2群比較のメタアナリシスの場合、「高品質」データのモデルは、TSD2[8]で紹介された標準モデルである.

$$\theta_{ik} = \mu_i + \delta_{i,1k} I_{\{k \neq 1\}}$$
 (8)

より低い品質のデータは、各試験が $\delta_{i,1k}$ に関する情報ではなく、バイアスのあるパラメータ $\delta_{i,1k}$ + β_i に関する情報を適用する.ここで、試験固有のバイアス項 β_i は、バイアスの期待値である平均 b_0 及び試験間分散 κ^2 である変量効果モデルの分布より導出されるものである.したがって、低品質の試験の場合:

$$\theta_{ik} = \mu_i + \left(\delta_{i,1k} + \beta_i\right) I_{\{k \neq 1\}}$$

$$\beta_i \sim N(b_0, \kappa^2)$$
(9)

 b_0 及び試験間分散 κ^2 の値は、外部データセットのベイズ流の分析、例えばメタアナリシスの集積 [4][65]等から得られ、これらの値は式(9)の事前分布に組み込むことができる.この分析では少なくとも 3つの推定値を生成する必要がある:バイアス β_i の期待値 b_0 、 b_0 の推定値の標準誤差,及びバイアスの試験間変動 κ^2 である.そして、これらの値は、新しいメタアナリシスでより低品質の試験対して治療効果を調整し、重みを小さくするように、バイアスのパラメータの事前分布を得るのに使うことができる. Weltonら[60]は、このバイアス補正法で必要となる仮定に関して述べている.この解析は、対象データセットの試験固有バイアスが、調整に使用される事前分布を与えるのに使用されるメタ疫学的データでのバイアスと交換可能と見なせるか、そして実際に、それらのバイアスが、意思決定の場面で、関係するステークホルダーにより交換可能とみなされるかどうかに大きく依存している.

本TSD3の執筆時点(2011年)では、メタ疫学研究からのデータ解析は、交換可能性の要件を恐らく満たしながらも、事前分布を得るためには、まだ利用可能ではなかった。それにもかかわらず、合理的なバイアス補正された分析は、「未調整の分析より、真のパラメータのより良く反映する可能性が高い」と(費用対効果評価において)判断される可能性がある。Weltonら[60]は、バイアス分布の特定の値のセッ

トに疑問がある場合でも、研究者は、より低品質の試験(潜在的に楽観的すぎる結果を有している)が存在しても、意思決定に影響を与えていないことを示すため、幾つかの感度分析を実行する可能性を示唆した. 広範囲に亘るメタ疫学的分析は、現在活発で興味のある研究領域である. バイアスの程度は、アウトカム指標の性質に依存し、(患者又は医師が報告した)主観的アウトカムではより大きく、「死亡」では事実上、バイアスが検出されないことは、すでに明らかである(極めて客観的である)[5]. 品質関連のバイアス(risk of bias)に関するより詳細な情報は公表することが求められ、また各領域で特定のアウトカムの種類や病状に対する事前分布のセットが必要となる.

原則として、同じタイプのバイアス補正は、新規性バイアス、スポンサーシップバイアス、又は小規模な試験バイアスのような、他のタイプのバイアスにも拡張され、あるいは無作為化比較試験と観察試験の混合へも拡張することが可能かもしれない. しかし、これらの拡張は、未だ実施されていない非常に大規模なメタ疫学データセットの詳細で広範な分析に依存している.

次に、バイアス分布の平均 b_0 及び分散 κ^2 を、興味のあるデータセット内で、内的に見積もることで、バイアス以外のパラメータにおける強い「交換可能性」の仮定に関連する困難(補足:バイアスのために他のパラメータにおける交換可能性が仮定できない困難)を取り除く方法を考える.

3.8.3 試験ネットワークでのバイアスの推定と調整

前述のバイアスのモデル(式(8),式(9))は、外部データに頼ることなく、内的にも見積もることができる.一連の試験を考えると、その中には品質が「高い」ものと「低い」ものがある.バイアスを内的に見積もる分析を、試験間でのバイアスの大きさ及びバラツキを知るために、常に利用することが可能であるが、バイアスのあるデータからいつも力を借りることができるとは限らない.例えば、治療法が2つしかない場合、分析によりバイアス分布は得られるが、真の治療効果に関しては判らない.このため、高品質のデータだけを参照することも可能である.

しかし、間接比較やネットワークでのデータ統合では、試験固有のバイアスの平均と分散が、各治療で同じであると仮定すると、1つの分析で、治療効果とバイアスを同時に推定し、それゆえ高品質と低品質の両方の試験を含むデータ全体に基づいた治療効果を推定し、バイアスを調整することが可能となる[61]. モデルは前節と正確に同じである.

$$\theta_{ik} = \mu_i + \left(\delta_{i,1k} + \beta_{ik} x_i\right) I_{\{k \neq 1\}}$$

ここで、試験iではバイアスがあると考えられ、それ以外の試験ではないと考えられる場合は、 $x_i=1$ となり、 β_{ik} は試験iの治療1群に対する治療k群の試験固有のバイアスである。Aがプラセボ又は標準治療であり、B、C、Dが全て実薬治療である場合、AB、AC、B0 及びAD試験に同じバイアス分布を適用できると期待することは合理的かもしれない。しかし、BC0 の、BD0 及びCD試験のバイアスモデルをどのようにコード化するかは、あまり明確ではない。実薬とプラセボ/標準治療を比較した試験と、実薬同士を比較した試験を区別にするかもしれない。平均バイアスが常に新しい治療に有利であると仮定すると、これは新規性バイアスのモデルになる[62]1. も[62]2 も[62]3 も[63]4 も[63]5 は、[63]5 は、[63]6 は、最初のタイプの試験は、バイアスの平均 [63]6 は、[63]7 の名は、最初のタイプの試験は、バイアスの平均 [63]7 の子が

 $N(b_1,\kappa^2)$ を有する正規分布に従うと仮定されるバイアス項を有する。実薬同士の比較試験は、最新の治療に有利と仮定するか、又はゼロに設定しうる異なるバイアスの平均 $b_2:\beta_{ik}\sim N(b_2,\kappa^2)$ を有すると仮定できる(詳細はDiasら[61]を参照)。この方法は、原則、臨床試験と観察研究の混合であるデータの統合を含むものに拡張できるが、これは未だ試みられていないようである。また、「内部」バイアスの任意の形式に拡張することもできる。Salantiら[62]は、がん試験での新規性バイアスの研究でこのモデルを適用した。特に興味深い応用は、「出版バイアス」の解釈の1つである「小規模な試験バイアス」である。この考えは、小さい試験ほどバイアスが大きくなるというものである。したがって、「真の」治療効果は、無限の大きさの試験で得られる効果として考えられる。これは、今度は、試験のバラツキに対する治療効果の回帰分析での切片として解釈される。Morenoら[63][66]は、このアプローチからのバイアス補正された推定値が、前向きに報告されたデータのレジスタに基づく単純なメタアナリシスで見出された結果に近似的に近づくことを示した。大規模なネットワークでは、「実薬同士の比較」試験でのバイアスの方向をどのようにコードするかに注意を払う必要がある。

3.8.2節で説明した方法のように、これらの方法は、半実験的なものと見なされるかもしれない.確かに応用に関する更なる経験が必要であり、提案されている内的バイアスの多くの形式間の関係性に関するさらなるメタ疫学的データが特に必要とされている[67]. しかし、それらはバイアス補正の合理的で妥当な方法を表しているように見え、データの品質が混在する状況で、バイアス補正を行わない場合よりも、優れている可能性が高い. 同時に、本方法は、本質的に「試験間」比較に関するメタ回帰である.試験の質のマーカーと効果の大きさとの間に「因果関係」があるという直接的なエビデンスはない(補足:間接的にはある可能性を示唆している. 試験規模が品質のマーカーになることがあるが、小規模試験では偶然効果が真値より大きくなる場合があり、効果が大きいと公表バイアスが生じやすい. ただし公表バイアスがなければ、小規模試験では偶然効果が真値より小さくなる場合も同様に生じる). したがって、小規模なデータセット(少数の試験)での利用を避けること、さらに結果が統計的に頑健(ロバスト)であり、結果が少数の試験に依存しないことを確立することが重要である(補足:試験数はある程度リッチでないとバイアスを正確に推定するのが難しい).

本節及び前節での基礎的なバイアスモデルは同じであるため、それらを併合させることは十分に実現可能だが、これも以前には行われていない.

3.8.4 専門家又はデータに基づいたバイアス分布の導出

本方法[59]は概念的には全てのバイアス補正方法の中で最も簡単であり、臨床試験と観察試験に同様に適用可能である。また、実行するのが最も困難で時間がかかるものでもある。1つの利点は、メタ回帰アプローチ(3.8.1節、3.8.3節)として試験数が不十分な場合でも使用できることであるが(詳細は原著[59]を参照)、本質的な考え方は次の通りである。各試験は、事前規定されたプロトコルを用いて、幾人かの独立した専門家によって検討される。このプロトコルは、一連の潜在的な内的及び外的バイアスを列挙し、各専門家はバイアス分布を構築するのに利用される情報を提供するよう求められる。考慮される内的バイアスの中には、(観察研究での)選択バイアス、非応答バイアス、摩擦バイアス (attrition bias)などがある。試験は内的及び外的バイアスの両方に悩まされることがある。このプロセスが完了すると、各評価者から各試験のバイアス情報が1つのバイアス分布に結合される。次いで、評価者の分布は数学

的に統合される. 原著では、バイアス分布の平均及び分散は、その試験で、効果的に、新しく調整された、治療効果の推定値を算出するために、元の試験の(治療効果の)推定及びその分散を統計的に組み合わせている. 最終段階は従来の統合であり、各試験からの調整された治療効果及びその分散を、標準的な2群比較のメタアナリシス、間接比較、又はNMAのためのデータ入力として扱う. その後、TSD2[8] (3.5節)の方法を、調整された試験固有の推定値に適用することができる.

この方法[59]は完全な形で実行するには相当の時間と手間が必要である.潜在的にバイアスのある試験の推定値を,バイアスに関する専門家の意見に基づいて調整された推定値で置き換えるという重要な考え方は,多くの状況で,問題となっているパラメータに対する全体の解析の感度と同等である,ある程度徹底して,実行ができる方法である.

参考文献

- [1] National Institute for health and Care Excellence. Guide to the methods of technology appraisal 2013.
- [2] 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
- [3] Higgins, J.P.T., Green, S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.0 [updated February 2008]. The Cochrane Collaboration, Wiley, Chichester; 2008.
- [4] Schulz, K.F., Chalmers, I., Hayes, R.J., Altman, D.G. Empirical Evidence of Bias. Dimensions of Methodological Quality Associated With Estimates of Treatment Effects in Controlled Trials. JAMA, J 1995; 273(5):408-412.
- [5] Wood, L., Egger, M., Gluud, L.L., Schulz, K., Juni, P., Altman, D. et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. British Medical Journal 2008; 336:601-605.
- [6] Higgins, J.P.T., Thompson, S.G., Spiegelhalter, D.J. A re-evaluation of random-effects meta-analysis. Journal of the Royal Statistical Society (A) 2009; 172:137-159.
- [7] 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 2012; available from http://www.nicedsu.org.uk
- [8] Dias, S., Welton, N.J., Sutton, A.J., Ades, A.E. NICE DSU Technical Support Document 2: A generalised linear modelling framework for pair-wise and network meta-analysis. 2011; last updated April 2012; available from http://www.nicedsu.org.uk
- [9] Borenstein, M., Hedges, L.V., Higgins, J.P.T., Rothstein, H.R. Introduction to Meta-Analysis. Wiley, Chichester; 2009.
- [10] Higgins, J.P.T., Thompson, S.G. Controlling the risk of spurious findings from metaregression. Statistics In Medicine 2004; 23:1663-1682.
- [11]Rothman, K.J., Greenland, S., Lash, T.L. Modern Epidemiology. 3 ed. Lippincott, Williams & Wilkins, Philadelphia; 2008.
- [12] Govan, L., Ades, A.E., Weir, C.J., Welton, N.J., Langhorne, P. Controlling ecological bias in evidence synthesis of trials reporting on collapsed and overlapping covariate categories. Statistics In Medicine 2010; 29:1340-1356.

- [13] Dominici, F. Combining contingency tables with missing dimensions. Biometrics 2000;56:546-553.
- [14] Welton, N.J., Johnstone, E.C., Munafo, M.R. A Cost-Effectiveness Analysis of Genetic Testing to Aid Treatment Choice for Smoking Cessation. Nicotine and Tobacco Research 2008; 10(1):231-240.
- [15] Glenny, A.M., Altman, D.G., Song, F., Sakarovitch, C., Deeks, J.J., D'Amico, R. et al. Indirect comparisons of competing interventions. Health Technology Assessment 2005; 9(26).
- [16] Lu, G., Welton, N.J., Higgins, J.P.T., White, I.R., Ades, A.E. Linear inference for Mixed Treatment Comparison Meta-analysis: A Two-stage Approach. Res Synth Method 2011;2:43-60.
- [17] Sidik, K., Jonkman, J.N. A comparison of heterogeneity variance estimators in combining results of studies. Statistics In Medicine 2007; 26:1964-1981.
- [18] Rucker, G., Schwarzer, G., Carpenter, J.R., Schumacher, M. Undue reliance on I2 in assessing heterogeneity may mislead. BMC Medical Research Methodology 2008; 8:79.
- [19] Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society (B) 2002; 64(4):583-616.
- [20] Higgins, J.P.T., Thompson, S.G. Quantifying heterogeneity in a meta-analysis. Statistics In Medicine 2002; 21:1539-1558.
- [21] Dakin, H., Fidler, C., Harper, C. Mixed Treatment Comparison Meta-Analysis Evaluating the Relative Efficacy of Nucleos(t)ides for Treatment of Nucleos(t)ide-Naïve Patients with Chronic Hepatitis B. Value in Health 2010; 13:934-945.
- [22] Spiegelhalter, D.J., Abrams, K.R., Myles, J. Bayesian approaches to clinical trials and Health-Care Evaluation. Wiley, New York; 2004.
- [23] Ades, A.E., Lu, G., Higgins, J.P.T. The interpretation of random effects meta-analysis in decision models. Med Decis Making, MED 2005; 25(6):646-654.
- [24] Welton, N.J., White, I., Lu, G., Higgins, J.P.T., Ades, A.E., Hilden, J. Correction: Interpretation of random effects meta-analysis in decision models. Med Decis Making, MED 2007; 27:212-214.
- [25] Thompson, S.G. Why sources of heterogeneity in meta-analyses should be investigated. British Medical Journal 1994; 309:1351-1355.
- [26] Thompson, S.G. Why and how sources of heterogeneity should be investigated. In: Egger M., Davey Smith G., Altman D., eds. Systematic reviews in health care: metaanalysis in context. 2nd ed. BMJ Books; London: 2001; 157-175.
- [27] Sterne, J.A.C., Bradburn, M.J., Egger, M. Meta-analysis in Stata. In: Egger M, Davey Smith G, Altman DG, eds. Systematic reviews in health care: Meta-analysis in context. BMJ Books; London: 2001; 347-369.
- [28] Marshall, E.C., Spiegelhalter, D.J. Approximate cross-validatory predictive checks in disease mapping models. Stat Med 2003; 22(10):1649-1660.
- [29] DuMouchel, W. Predictive cross-validation of Bayesian meta-analyses. In: Bernardo J.M., Berger J.O., Dawid A.P., Smith A.F.M., eds. Bayesian Statistics 5. Oxford University Press; Oxford: 1996; 107-127.
- [30] Egger, M., Davey-Smith, G. Misleading meta-analysis. British Medical Journal 1995;310:752-754.
- [31] Higgins, J.P.T., Spiegelhalter, D.J. Being sceptical about meta-analyses: a Bayesian perspective on magnesium trials in myocardial infarction. International Journal Of Epidemiology 2002; 31(1):96-104.
- [32]Li, J., Zhang, Q., Zhang, M., Egger, M. Intravenous magnesium for acute myocardial infarction. Cochrane

- Database of Systematic Reviews 2007; 2009, Issue 4:Art. No.: CD002755.
- [33] Raiffa, H., Schlaiffer, R. Applied statistical decision theory. Wiley Classics Library ed. Wiley Interscience, New York; 1967.
- [34] Dias, S., Welton, N.J., Caldwell, D.M., Ades, A.E. Checking consistency in mixed treatment comparison meta analysis. Statistics In Medicine 2010; 29:932-944.
- [35]Sutton AJ. Meta-analysis methods for combining information from different sources evaluating health interventions. Thesis/Dissertation: University of Leicester, UK; 2002.
- [36]Cooper, N.J., Sutton, A.J., Morris, D., Ades, A.E., Welton, N.J. Adressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. Statistics In Medicine 2009; 28:1861-1881.
- [37] Nixon, R., Bansback, N., Brennan, A. Using mixed treatment comparisons and metaregression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. Statistics In Medicine 2007; 26(6):1237-1254.
- [38] Berkey, C.S., Hoaglin, D.C., Mosteller, F., Colditz, G.A. A random effects regression model for meta-analysis. Statistics In Medicine 1995; 14:395-411.
- [39] National Institute for health and Care Excellence. Certolizumab pegol for the treatment of rheumatoid arthritis. 2010; TA186. NICE technology appraisal guidance.
- [40]McIntosh, M.W. The population risk as an explanatory variable in research synthesis of clinical trials. Statistics In Medicine 1996; 15:1713-1728.
- [41] Thompson, S.G., Smith, T.C., Sharp, S.J. Investigating underlying risk as a source of heterogeneity in metaanalysis. Statistics In Medicine 1997; 16:2741-2758.
- [42] Stewart, L.A., Clarke, M.J. Practical methodology of meta-analyses (overviews) using updated individual patient data. Statistics In Medicine 1995; 14:2057-2079.
- [43]Riley, R.D., Lambert, P.C., Abo-Zaid, G. Meta-analysis of individual participant data: rationale, conduct, and reporting. British Medical Journal 2010; 340:c221.
- [44] Lambert, P.C., Sutton, A.J., Abrams, K.R., Jones, D.R. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. Journal of Clinical Epidemiology 2002; 55:86-94.
- [45] Berlin, J.A., Santanna, J., Schmid, C.H., Szczech, L.A., Feldman, H.I. Individual patientversus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. Statistics In Medicine 2002; 21:371-387.
- [46] Simmonds, M.C., Higgins, J.P.T., Stewart, L.A., Tierney, J.F., Clarke, M.J., Thompson, S.G. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. Clinical Trials 2005; 2:209-217.
- [47]Simmonds, M.C., Higgins, J.P.T. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. Statistics In Medicine 2007; 26:2982-2999.
- [48] Higgins, J.P.T., Whitehead, A., Turner, R.M., Omar, R.Z., Thompson, S.G. Metaanalysis of continuous outcome data from individual patients. Statistics In Medicine 2001; 20:2219-2241.
- [49] Goldstein, H., Yang, M., Omar, R.Z., Turner, R.M., Thompson, S.G. Meta-analysis using multilevel models with an application to the study of class size effects. Applied Statistics 2000; 49:399-412.

- [50] Turner, R.M., Omar, R.Z., Yang, M., Goldstein, H., Thompson, S.G. A multilevel model framework for metaanalysis of clinical trials with binary outcomes. Statistics In Medicine 2000; 19:3417-3432.
- [51] Tudor Smith, C., Williamson, P.R., Marson, A.G. Investigating heterogeneity in an individual patient data metaanalysis of time to event outcomes. Statistics In Medicine 2005; 24:1307-1319.
- [52] Whitehead, A., Omar, R.Z., Higgins, J.P.T., Savaluny, E., Turner, R.M., Thompson, S.G. Meta-analysis of ordinal outcomes using individual patient data. Statistics In Medicine 2001; 20:2243-2260.
- [53] Tudor Smith, C., Marson, A.G., Chadwick, D.W., Williamson, P.R. Multiple treatment comparisons in epilepsy monotherapy trials. Trials 2007; 8:34.
- [54] Riley, R.D., Steyerberg, E.W. Meta-analysis of a binary outcome using individual participant data and aggregate data. Res Synth Method 2010; 1:17.
- [55]Riley, R.D., Simmonds, M.C., Look, M.P. Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. Journal of Clinical Epidemiology 2007; 60:431-439.
- [56] Sutton, A.J., Kendrick, D., Coupland, C.A.C. Meta-analysis of individual- and aggregate-level data. Statistics In Medicine 2008; 27:651-669.
- [57] Riley, R.D., Lambert, P.C., Staessen, J.A., Wang, J., Gueyffier, F., Boutitie, F. Metaanalysis of continuous outcomes combining individual patient data and aggregate data. Statistics In Medicine 2008; 27:1870-1893.
- [58] Saramago, P., Sutton, A.J., Cooper, N., Manca, A. Synthesizing effectiveness evidence from aggregate-and individual-patient level data for use in cost-effectiveness modelling. Presented at HESG Winter Conference, Centre for Health Economics, University of York. 2011.
- [59] Turner, R.M., Spiegelhalter, D.J., Smith, G.C.S., Thompson, S.G. Bias modelling in evidence synthesis. Journal of the Royal Statistical Society (A) 2009; 172:21-47.
- [60] Welton, N.J., Ades, A.E., Carlin, J.B., Altman, D.G., Sterne, J.A.C. Models for potentially biased evidence in meta-analysis using empirically based priors. Journal of the Royal Statistical Society (A) 2009; 172(1):119-136.
- [61]Dias, S., Welton, N.J., Marinho, V.C.C., Salanti, G., Higgins, J.P.T., Ades, A.E. Estimation and adjustment of bias in randomised evidence by using Mixed Treatment Comparison Meta-analysis. Journal of the Royal Statistical Society (A) 2010; 173(3):613-629.
- [62] Salanti, G., Dias, S., Welton, N.J., Ades, A.E., Golfinopoulos, V., Kyrgiou, M. et al. Evaluating novel agent effects in multiple treatments meta-regression. Statistics In Medicine 2010; 29:2369-2383.
- [63] Moreno, S.G., Sutton, A.J., Turner, E.H., Abrams, K.R., Cooper, N.J., Palmer, T.M. et al. Novel methods to deal with publication biases: secondary analysis of antidepressant trials in the FDA trial registry database and related journal publications. BMJ 2009; 339:b2981.
- [64] Thompson, S., Ekelund, U., Jebb, S., Karin Lindroos, A., Mander, A., Sharp, S. et al. A proposed method of bias adjustment for meta-analyses of published observational studies. Internation Journal of Epidemiology 2010; doi: 10.1093/ije/dyq248.
- [65] Savovic, J., Harris, R.J., Wood, L., Beynon, R., Altman, D., Als-Nielsen, B. et al. Development of a combined database for meta-epidemiological research. Res Synth Method 2010; 1:212-225.
- [66] Moreno, S.G., Sutton, A.J., Ades, A.E., Stanley, T.D., Abrams, K.R., Peters, J.L. et al. Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. BMC Medical Research

Methodology 2009; 9(2):doi:10.1186/1471-2288-9-2.

[67]Dias, S., Welton, N.J., Ades, A.E. Study designs to detect sponsorship and other biases in systematic reviews. Journal of Clinical Epidemiology 2010; 63:587-588.

4 無作為化比較試験に基づくエビデンスのネットワークにおける不一致性 (TSD4)

NMA (Network meta-analysis) における不一致性(Inconsistency)とは、ネットワークにおける各治療間の直接比較から得られるエビデンスと間接比較から得られるエビデンスの不整合のことである. NICE の公開する TSD4[1]は無作為化比較試験に基づくエビデンスのネットワークにおける不一致性に関する技術的文書であり、不一致性の検出方法や不一致性への対応について解説している.

例として、3つの治療 A,B,C をそれぞれ対比較した AB 試験、AC 試験、BC 試験について固定効果モデルの NMA を考えてみる。各試験 i において、実際に比較された治療にかかわらず治療 B に対する治療 C の真の相対治療効果 δ_{iBC} はどれも同じであり、 $\delta_{iBC}=d_{BC}$ という仮定から、直接比較の相対治療効果と間接比較の相対治療効果が等しいという一致性の等式(consistency equations) が成り立つことを前提に NMA は行われる(TSD2 参照)。一致性の等式が成り立たないとき、すなわち BC 試験から推定した相対治療効果 d_{BC} と、AB 試験と AC 試験による間接比較から推定した相対治療効果 $d_{AC}-d_{AB}$ が異なるときに不一致性がみられる。

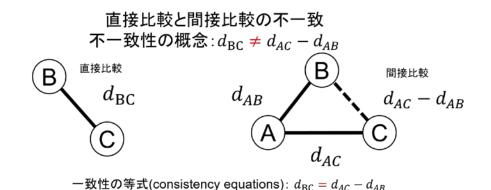


図 4-1 不一致性の概念図

以下は TSD4 に記載されている推奨事項をまとめたものであり、本章ではこれらについて解説する.

- 潜在的な不一致性を把握するため、ネットワークダイアグラムを作成・調査すべきである.
- 不一致性を検出する方法としては、単純なネットワークの場合は、ネットワークの特定部分に着目して直接効果と間接効果に差があるかの検定を行う Bucher 法とその拡張法がシンプルで強力である。複雑なネットワークの場合は、Bucher 法とその拡張法を繰り返し用いるときは検定の多重性に注意すべきであり、一致性を仮定した通常のモデル(consistency model)と一致性を仮定しないモデル(inconsistency model: Unrelated mean effects (UME) model)の比較や Node-splitting 法が有用である。

- 不一致性は異質性と同様に、効果修飾因子(effect-modifier)によって引き起こされるため、不一致性の確認は、異質性の原因を調整した後に実施すべきである.
- 不一致性が検出された場合,潜在的な効果修飾因子の存在に関して,臨床疫学の観点からエビデンスのネットワーク全体を再考すべきである.

4.1 エビデンスのループ

TSD4 では不一致性の評価の際には、ネットワークを図示したネットワークダイアグラムを調べることを強く推奨している。ネットワークダイアグラムでは、各ノード(節点)が治療方法、ノードを結ぶ辺(edge)が2つの治療の比較試験の存在を表している。図 42 a)について着目すると、治療 A-B-C のループ1つで表されており(以下、ループ ABC と表記)、TSD2 の一致性を仮定したモデルでは、治療 Aを NMA の参照治療とすると、基本パラメータとして dABと dACがあり、dBCは他の2つのパラメータから一致性の等式により求まる。ネットワークの情報としては、3つのデータソースと2つの独立したパラメータと捉えることができ、潜在的な不一致性として自由度が1つあるとみなせる。TSD4ではこの自由度をInconsistency Degrees of Freedom (ICDF) = "N-基本パラメータ数"と定義している。ここでNは対比較の数、基本パラメータ数は"治療数(nt)-1"を示し、ネットワークの試験全てが2群比較試験のときには、ICDFはN-(nt-1)で求まる。ICDFは独立なループの数と同じであり、2群比較試験のネットワークに一つの独立なループが加わるとICDFが1増加する。例えば図42c)の左側のABCDの四角形にはループABC、ループBCD、ループABDCの3つのループがあり、このうち2つがわかると残りの辺がわかるため、独立なループは2つ、つまり、ICDFは2である。右側のループEFGは単一の独立なループであり、左側のABCDのICDFと合計すると3となり、全体のICDFの計算結果とも一致する。

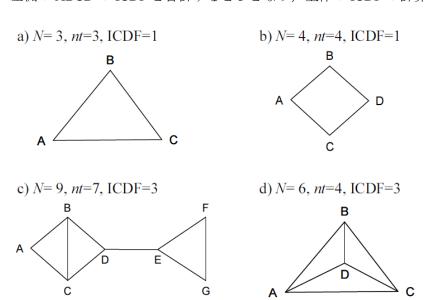


図 4-2 ネットワークダイアグラムの例

「出典: NICE DSU TSD4]

4.2 多群試験の影響

ネットワークに多群試験が含まれると不一致性の定義は複雑になる.図42a)が3群試験だけで構成されている場合は3辺全てのエビデンスを提供することになるので不一致性は存在せず,あるとしたら試験間の異質性だけとなる.ICDFは独立なループの数に相当するので、1つのループが多群試験だけから形成される場合は、独立なループとしては数えない。AB,AC,BC,ABCの試験があるときを考えてみると、この場合は主に2つの問題がある.1つ目はABC試験からのエビデンスとAB試験のエビデンスとの潜在的な不一致性を考慮するかという問題である。2つ目の問題はパラメータ化の選択である。ABC試験における独立な情報は3つの比較のうち2つだけであり、どの比較をパラメータにするかは異質性の推定と不一致性の検出の両方に影響がある。次節で紹介する不一致性を検出するBucher法では、多群試験を除外するという方法がTSD4で提案されている。満足できる方法ではないものの、多くの試験が2群であろうことから、不一致性の検出に重大な影響があることは少ないだろうとのことだ。

4.3 不一致性の評価方法

4.3.1 Bucher 法: ひとつのループがあるネットワークへの適用

不一致性の検定方法として最も単純で強力なものとしてTSD4ではBucherらの提案する方法(以下, Bucher法) [2] を推奨している.この方法は2段階となっており、1段階目で対比較のエビデンスをそれぞれ統合し、2段階目で直接比較と間接比較から得られるエビデンスが異なっているかを検定する.

図 4-2 a)のループ ABC のネットワークを例に考える.BC の直接比較 $\hat{\mathbf{d}}_{BC}^{pir}$ を AB と AC から成る間接比較 $\hat{\mathbf{d}}_{BC}^{Imd}$ (= $\hat{\mathbf{d}}_{AC}^{pir}$ – $\hat{\mathbf{d}}_{AB}^{pir}$) と比較する.直接比較の推定値は独立なので,分散は,

$$\operatorname{Var}(\hat{\mathbf{d}}_{BC}^{Ind}) = \operatorname{Var}(\hat{\mathbf{d}}_{AC}^{Dir}) + \operatorname{Var}(\hat{\mathbf{d}}_{AB}^{Dir})$$

となり、不一致性の程度をωとすると次式で表せる.

$$\widehat{\omega}_{BC} = \widehat{\mathbf{d}}_{BC}^{Dir} - \widehat{\mathbf{d}}_{BC}^{Ind}$$

$$\operatorname{Var} \left(\widehat{\boldsymbol{\omega}}_{BC} \right) = \operatorname{Var} \left(\widehat{\boldsymbol{d}}_{BC}^{Dir} \right) + \operatorname{Var} \left(\widehat{\boldsymbol{d}}_{BC}^{Ind} \right) = \operatorname{Var} \left(\widehat{\boldsymbol{d}}_{BC}^{Dir} \right) + \operatorname{Var} \left(\widehat{\boldsymbol{d}}_{AC}^{Dir} \right) + \operatorname{Var} \left(\widehat{\boldsymbol{d}}_{AC}^{Dir} \right)$$

ここで,不一致性がないという帰無仮説を $\mathbf{Z}_{\mathrm{BC}} = \frac{\hat{\omega}_{BC}}{\sqrt{Var(\hat{\omega}_{BC})}}$ が標準正規分布に従うことに基づき検定する.

前述のとおり、Bucher 法に3群試験を含めてしまうと、3群試験は内部で整合性がとれているため、 検出力が低下することになる。また、不一致性の分散は全ての比較の分散の合計であることから、ルー プ中の治療の数が増えると分散が大きくなり、不一致性は検出されにくくなる。TSD4 ではこの場合 3 群試験を除外して検定することを提案している。

適用例: HIV 抑制に関する3つ治療のメタアナリシスは,図4-2a)のネットワークと同じ構造であり,

直接比較の推定値は以下の対数オッズ比で示された[3].

表 4-1 HIV 抑制 に関する 3 つ治療のメタアナリシス

[出典: NICE DSU TSD4]

ここで、治療 C と B の相対治療効果の間接比較の推定値は $\hat{\mathbf{d}}_{BC}^{Ind} = 1.42 - 2.79 = -1.37$ 、 $Var(\hat{\mathbf{d}}_{BC}^{Ind}) = 0.56^2 + 0.34^2 = 0.429$ である。直接推定 $\hat{\mathbf{d}}_{BC}^{Dir} = 0.47$ と比較すると、不一致性の程度の推定値 $\hat{\boldsymbol{\omega}}_{BC} = 0.47 - (-1.37) = 1.84$ と $Var(\hat{\boldsymbol{\omega}}_{BC}) = 0.10^2 + 0.429$ が得られる。従って、 $Z_{BC} = \frac{1.84}{\sqrt{0.439}} = 2.78$ となり、不一致性が検出された(p-value < 0.01)。

統計的交互作用の検出と同様に、不一致の検出は治療効果を検出するよりも多くのデータを要求する. そのため帰無仮説を棄却できなかったとしても、検出力不足が理由の可能性があることに留意されたい.

4.3.2 Bucher 法の拡張:複数のループがあるネットワークへの適用

複数のループがある際に不一致性の分析をそれぞれ独立に行うことができるパターンとして、図 4-2c) のネットワーク構造がある. 左側のループ ABCD と右側のループ EFG の 2 つにわけて考えることで、例えばループ EFG では EF について Bucher 法を適用し、ループ ABCD の BC については Bucher 法を分散の逆数で重み付けする逆分散法と組み合わせてカイ 2 乗検定に拡張して適用できる. 本文書では割愛するが、TSD4 では数値例も記載されている.

4.4 複雑なネットワークにおける不一致性の評価方法

図 4-2 c)と d)はどちらも ICDF は 3 であるものの、d)には、3 つの治療によるループ(ACD, BCD, ABD, ABC)が 4 つ、4 つの治療によるループ(ABCD, ACDB, CABD)が 3 つあり、これらのループが独立ではないという点が c)と異なっており、これは 3 つの潜在的な不一致性に対しての独立な検定を構築することができないことを示している。Bucher 法を 7 つのループそれぞれに適用することを考えてみると、ループの数と検定の数はネットワークが持つ不一致性の数を超えてしまう。このようなネットワークをTSD4 では複雑なネットワークと呼んでいる。

ある文献では 133 個のループのうち 6 つのループで不一致性が検出されたが、133 回の検定を行っているため、この結果は多重性による偶然の可能性があると著者は述べている[4]. これに対して TSD4 では、この 133 個のループは独立ではないため、この結論に対して疑問であると述べている。もしネットワークの全てのループについて検定する場合、検定の多重性を考慮して有意水準を調整する必要がある。しかし、これらの検定は独立ではなく、厳密に調整した有意水準に修正する計算は複雑となる。さらに、

複雑なネットワークでは間接比較に用いられる治療やループが多くなるため、不一致性が検出されにくくなることも考えられる. そのため、Bucher 法を繰り返すことは比較的容易ではあるものの、TSD4では次に示す方法を推奨している.

4.4.1 一致性を仮定しないモデル(Inconsistency model: Unrelated mean effects model)

複雑なネットワークでは、Bucher法を繰り返し適用する代わりに、TSD2で提示された一致性を仮定したモデル(consistency model)と一致性を仮定しないモデル(inconsistency model)を比較することをTSD4では提案している。ここで提案されている一致性を仮定しないモデルとは、利用可能な対比較のそれぞれが無関係であるとして推定するモデル(Unrelated mean effects model; UME model)である。なお、TSD4ではこのモデルを単にinconsistency modelと表記しているが、本文書では他の一致性を仮定しないモデルと区別するために、TSDの著者らが執筆している書籍[5]と同様にUME modelと表記する。

一致性を仮定したモデルでは、nt個の治療A,B,C…を伴うネットワークにおいてnt-1個の基本パラメータ d_{AB} 、 d_{AC} …を定義する。つまり参照治療として選択された治療Aと比較したときの各治療の効果を推定することになる。他の対比較は全て派生したパラメータであり、それらは一致性の等式により基本パラメータの関数として定義できる。例えば図4-2b)のネットワークでは4つの対比較があり、一致性を仮定したモデルでは3つの相対治療効果 d_{AB} 、 d_{AC} 、 d_{AD} を基本パラメータとし、 d_{CD} は d_{AD} - d_{AC} で定義される。一方UME modelでは、パラメータ間の関係に一致性の等式の仮定をおかず、4つの対比較におけるエビデンスから d_{AB} 、 d_{AC} 、 d_{BD} 、 d_{CD} 04つの相対治療効果を推定する。

試験がM個あるとすると、変量効果モデルでは、ある試験iでの治療効果 $\delta_{i,XY}$ は以下の正規分布に従うと仮定する。UME modelでは、平均治療効果のそれぞれで共通の分散 σ^2 を共有する別々のパラメータとして扱われる。

$$\delta_{i,XY} \sim N(d_{XY}, \sigma^2)$$
 for $i = 1, ..., M$

多群試験が含まれるときUME modelでは、複数の対比のうちどこに着目するかによって異なるパラメータ化を行う. 例えば、ABC試験では、Aを対照群としたABとACの相対治療効果をパラメータとすることもできるし、Bを対照群としたABとBCの相対治療効果、Cを対照群としたACとBCの相対治療効果とすることも選択できる. パラメータ化の選択はパラメータ推定とその不一致性の検定に影響を与えることに注意されたい.

この方法の問題点は、試験間の変動を推定するための十分なデータがあることが稀だということである。試験間変動について無情報事前分布を使うことは、不一致性をみえにくくしてしまう可能性が高い。 TSD4ではUME modelの解析を開始する前に、一致性を仮定したモデルから試験間変動の事後分布を精査することを推奨している。データが現実的な値にならない場合は、専門家の意見や疫学データに基づいた情報をもった事前分布を使うことを考慮すべきである。

4.4.2 不一致性を評価するその他の方法

● 不一致性を表現するモデル

上で述べたUME modelとは別のパラメータによる一致性を仮定しないモデルも考えられる. 例えば、図4-2(d)における4つの治療ネットワークでは、一致性を仮定したモデルにおける参照治療をAとした3つの基本パラメータ d_{AB} , d_{AC} , d_{AD} に対して、UME modelでは6つのパラメータ d_{AB} , d_{AC} , d_{AD} , d_{BD} , d_{CD} をもつが、このモデルでは d_{BC} , d_{BD} , d_{CD} の代わりに以下の ω_{BC} , ω_{BD} , ω_{CD} を用いて不一致性を表現する.

$$\begin{cases} \omega_{BC} = d_{BC} - (d_{AC} - d_{AB}) \\ \omega_{BD} = d_{BD} - (d_{AD} - d_{AB}) \\ \omega_{CD} = d_{CD} - (d_{AD} - d_{AC}) \end{cases}$$

このパラメータは一致性の仮定が成り立つときには0であり、不一致性の尺度をあらわすものとして incoherence variance[6], inconsistency variance[7]として提案されてはいるものの、TSD4ではこれらを不一 致性の指標としては推奨していない。ICDFが大きくないときでも、 ω は非常に広い信用区間となり、単 一ループ(ICDF=1)では σ_{ω} の推定値を得ることができない場合があるからである。それにも関わらず、単一ループであっても、このモデルを用いた不一致性の評価は数多く報告されており、ICDFを考慮しているようにはみえないとも言及している。この課題におけるさらなる議論については、Salanti (2008) [8] を参照されたい。

● Node-Splitting法

Node-Splitting法[9]は、相対治療効果d_{XY}を2つの要素: XYデータ全て(XY試験, XYZ試験, WXY試験 など)からなる直接比較と、残りのデータによる間接比較に分離してベイズ流の枠組みでパラメータを推定する方法である。直接比較と間接比較のパラメータの事後分布を図示することで直感的に違いが判断でき、また高い検出力を有する頑健(ロバスト)な方法であることから、複雑なネットワークにおける不一致性の評価のオプションとしてTSD4では推奨している。

Node-Splitting法はパラメータ化が容易ではないとも記載されており、TSD4では実行例やWinBUGSプログラムは提示されていない。ただし、TSD4の著者らはNode-splitting法を含む不一致性の検討を報告しており、その付録にはRとWinBUGSによるプログラムがついている[9]。また、Rのgemtcパッケージには適用可能な箇所に自動的にNode-Splitting法を適用する方法が実装されている[10]。なお、本TFでは統計解析ソフトSASによる実装も試みており、次節の例ではSASによるNode-splitting法の結果を提示した。

● ネットヒートプロット

TSD4公開以降も不一致性を評価する方法は研究されている. 例えば, ある比較において直接比較を除いた場合に不一致性の程度が減少するほど暖色, 増加するほど寒色になるヒートマップをネットヒートプロット(net heat plot)としてKrahnらが提案している[11].

4.4.3 複雑なネットワークにおける不一致性評価の事例

TSD4に示されている事例のうち、急性心筋梗塞に対する血栓溶解療法の事例を紹介する. 50試験のデータセットには以下の治療群の30日間または35日間での死亡数と患者数が含まれる.

治療法(略称,番号): ストレプトキナーゼ (SK,1), アルテプラーゼ (t-PA,2), アルテプラーゼ急速 静注 (Acc t-PA,3), ストレプトキナーゼ・アルテプラーゼ併用 (SK+t-PA,4), レテプラーゼ (r-PA,5), テノクテプラーゼ (TNK,6), 経皮経管冠動脈形成術 (PTCA,7), ウロキナーゼ (UK,8), アニストレプ チラーゼ (ASPAC,9)

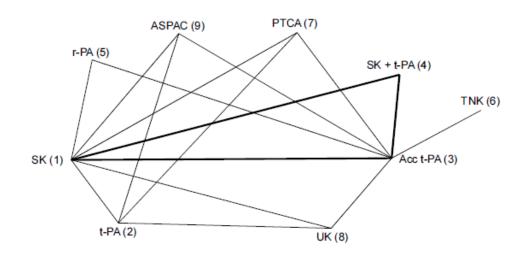


図 4-3 血栓溶解療法のネットワーク

[出典: NICE DSU TSD4]

合計9つの治療と16個の対比較にみえるが、図4-3の太線で強調されているループは3群試験のみで作られているため、ICDFは2群試験のみの場合の計算式の16-(9-1)からさらに1を引き7となる。またAcc t-PAと SK+t-PAの比較は3群試験のみで行われているため不一致性を考慮する必要はない。UME modelで15個の相対治療効果を固定効果として推定し、一致性を仮定したモデルによる推定値と比較した結果を表 4-2に示す。なお、一致性を仮定したモデルではSKを参照治療として、SKに対する8つの相対治療効果を基本パラメータとして推定し、残りの相対治療効果は一致性の等式から得ている。

表 4-2 一致性を仮定したモデルと UME model の結果 (X に対する Y の対数オッズ比)

treatr	nents	一致	性を仮定し	たモデル	UME model		
X	Y	mean	sd	95%信用区間	mean	sd	95%信用区間
SK(1)	t-PA(2)	0.002	0.030	(-0.06, 0.06)	-0.004	0.030	(-0.06, 0.06)
SK(1)	Acc t-PA(3)	-0.177	0.043	(-0.26, -0.09)	-0.158	0.049	(-0.25, -0.06)
SK(1)	SK+t-PA(4)	-0.049	0.046	(-0.14, 0.04)	-0.044	0.047	(-0.14, 0.05)
SK(1)	r-PA(5)	-0.124	0.060	(-0.24, -0.01)	-0.060	0.089	(-0.23, 0.11)
SK(1)	PTCA(7)	-0.476	0.101	(-0.67, -0.28)	-0.665	0.185	(-1.03, -0.31)
SK(1)	UK(8)	-0.203	0.221	(-0.64, 0.23)	-0.369	0.518	(-1.41, 0.63)
SK(1)	ASPAC(9)	0.016	0.037	(-0.06, 0.09)	0.005	0.037	(-0.07, 0.08)
t-PA(2)	PTCA(7)	-0.478	0.104	(-0.68, -0.27)	-0.544	0.417	(-1.38, 0.25)
t-PA(2)	UK(8)	-0.206	0.221	(-0.64, 0.23)	-0.294	0.347	(-0.99, 0.37)
t-PA(2)	ASPAC(9)	0.013	0.037	(-0.06, 0.09)	-0.290	0.361	(-1.01, 0.41)
Acc t-PA(3)	r-PA(5)	0.054	0.055	(-0.05, 0.16)	0.019	0.066	(-0.11, 0.15)
Acc t-PA(3)	TNK(6)	0.005	0.064	(-0.12, 0.13)	0.006	0.064	(-0.12, 0.13)
Acc t-PA(3)	PTCA(7)	-0.298	0.098	(-0.49, -0.11)	-0.216	0.119	(-0.45, 0.02)
Acc t-PA(3)	UK(8)	-0.026	0.221	(-0.45, 0.41)	0.146	0.358	(-0.54, 0.86)
Acc t-PA(3)	ASPAC(9)	0.193	0.056	(0.08, 0.30)	1.405	0.417	(0.63, 2.27)
残差逸脱原	隻 resdev	105.9			99.7		
p.	D	58			65		
Di	IC	163.9			164.7		

一致性を仮定したモデル: 固定効果モデル, 5000回の Burn-in, 2-chains, 繰り返し回数 50,000回

UME model: 固定効果モデル, 20000 回の Burn-in, 3-chains, 繰り返し回数 50,000 回

[出典: NICE DSU TSD4]

UME modelは一致性を仮定したモデルよりも残差逸脱度が小さくデータにあてはまっているようにみえるものの、DICは同程度である.二つのモデルの各データポイントにおける残差逸脱度の事後平均の散布図(図 4-4)をみると、4つのポイントでUME modelは一致性を仮定したモデルよりも低い値を示しており、一致性を仮定したモデルはあまりあてはまっていないことがわかる.これらのポイントはAcc t-PAとASPACを比較する試験番号44および45である.これらの対数オッズ比の95%信用区間が重なっていないこととあわせると不一致性が存在するとみなせる.不一致性がネットワークの特定の箇所にみられる場合、全体のネットワークを再検討し、試験の選択基準や効果修飾因子の違いを疑うことになる.

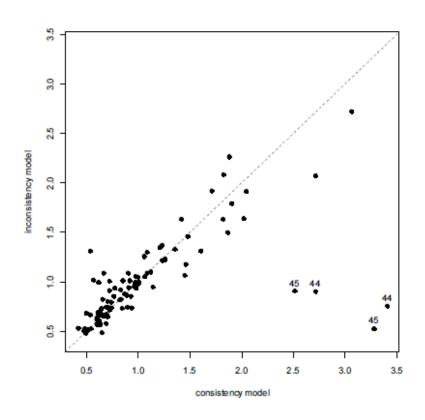


図 4-4 Residual deviance plot

[出典: NICE DSU TSD4]

TSD4付録のUME modelのWinBUGSコードに補足を加えたものを以下に示す. 各試験のデータと初期値については割愛するため、TSD4の付録を参照されたい. なお、変量効果モデルについては別事例に対しての適用例としてTSD4に記載されている. このWinBUGSコードは多群試験のパラメータ選択は第1群を参照群とした相対治療効果であるため、別のパラメータを選択する際には注意されたい.

```
# Binomial likelihood, logit link, inconsistency model
# Fixed effects model
                                                         # *** PROGRAM STARTS
model{
                                                        #試験のループ処理
for(i in 1:ns){
    mu[i] \sim dnorm(0,.0001)
                                                        # 試験のベースラインの事前分布
                                                        # 治療群のループ処理
    for (k in 1:na[i]) {
                                                        #二項尤度
        r[i,k] \sim dbin(p[i,k],n[i,k])
                                                                 #ここで直接比較の相対治療効果のパラメータを用いる
        logit(p[i,\!k]) <- mu[i] + \underline{\textbf{d[t[i,\!1],\!t[i,\!k]]}}
                                                        # 分子の期待値
rhat[i,k] \leftarrow p[i,k] * n[i,k]
```

```
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))</td>
      #残差逸脱度

      + (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))

      }
      | この試験の残差逸脱度

      | totresdev[i] <- sum(dev[i,1:na[i]])</td>
      # この試験の残差逸脱度

      | totresdev<- sum(resdev[])</td>
      # ネットワーク全体の残差逸脱度

      | for (k in 1:nt) { d[k,k] <- 0 }</td>
      # 同一治療の相対治療効果を0にセット

      | for (c in 1:(nt-1)) {
      # 平均治療効果についての事前分布

      | for (k in (c+1):nt) { d[c,k] ~ dnorm(0,0001) }
      | # **** PROGRAM ENDS pDの算出にはDICツールを使用
```

また、Node-splitting法によるAcc t-PAとASPACの比較における不一致性評価を論文[9]の表示例と同様に、Acc t-PAに対するASPACの対数オッズ比の直接比較と間接比較の相対治療効果Direct、Indirectの事後分布を図 4-5にプロットし、不一致の程度 $\hat{\omega}$ = Direct - Indirect、事後確率prob=Pr($\hat{\omega}$ > 0)、ベイジアンp-value p=2*min(prob,1-prob)を表 4-3に示した。これらの結果からも直接比較と間接比較が明らかに異なることがわかり、不一致性が検出された。Node-splitting法では各比較についてそれぞれ評価するため、UME modelのようにネットワーク全体のあてはまりを評価することはできない。なお、用いたSASプログラムのコードはSASユーザー総会で公開している[12]。

表 4-3 Node-splitting 法の結果(対数オッズ比)

Acc t-PA(3) vs ASPAC(9)	mean	SD	95%信用区間	$Pr(\widehat{\omega} > 0)$	p-value
Direct	1.41	0.42	(0.63, 2.27)		
Indirect	0.16	0.06	(0.05, 0.28)		
ω̂ (Direct-Indirect)	1.24	0.42	(0.45, 2.12)	0.9994	0.0012

Burn-in 10,000回, 繰り返し回数50,000回

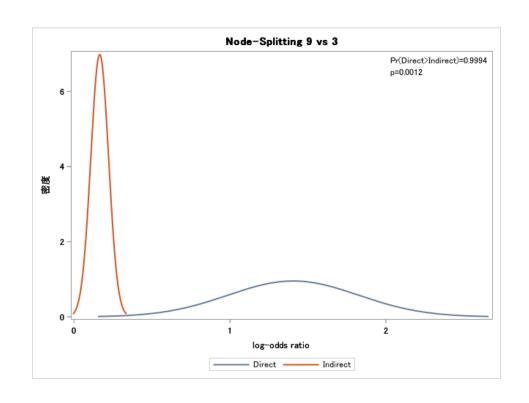


図 4-5 Node-splitting 法による相対治療効果の事後分布の比較

4.5 不一致性への対応, 異質性との関係

不一致性への対応と異質性との関係についてTSD4の内容を以下に要約する.

- 不一致性は効果修飾因子と治療比較に関係している異質性の特別な場合であることから、試験間の異質性は不一致性のリスクを増加させるが、分散の増大によって不一致の検出力は低くなる.
- 間接比較でバイアスを引き起こしている可能性のあるメカニズムは、対比較のメタアナリシスの異質性を引き起こしているものと同様と考えられる。そのため、間接比較に基づく結論が妥当であることを確認するために、直接比較のエビデンスについても確認しなければならない。直接比較におけるバイアスは、間接比較の推定値に引き継がれる。例えば直接比較が異なる患者集団で行われた試験に基づく場合に治療の効果修飾因子が存在すると、間接比較にはそれに起因するバイアスが生じうる。効果修飾因子の分布がたまたま不均衡であることに起因する不一致性は、各比較の試験数が増えるにつれて発生しにくくなるはずである。
- 試験間の異質性を避けるためのあらゆるステップは、誤った結論を導くリスクを減らすのに効果的である. 例えば、典型的な効果修飾因子である年齢、ベースラインの重症度、前治療がそれぞれ交絡しているかもしれない. ネットワークメタアナリスと不一致性の評価の前に各試験を吟味し、潜在的な交絡因子を確認し、潜在的なバイアスの補正とメタ回帰(TSD3参照)を検討すべきである.
- 異質性の評価は不一致のリスクがあることを注意喚起でき、 I^2 またはカイ2乗検定による異質性の検定

が利用できる(TSD3 参照). また、試験間の標準偏差の事後分布の要約指標を平均治療効果の大きさと比較することも有用かもしれない. さらにベースラインの試験間変動も指標になる. 参照群全てで同様なイベント率、ハザードをもつ場合は、試験集団が相対的に同質で、治療効果にほとんど異質性がないことになる. ベースラインが高い異質性を示したとしても、相対治療効果に異質性があるということを意味してはいないが、少なくとも異質性のリスクがあるといえる. 治療効果の評価尺度の選択(リスク差、logit、c-loglog など)によっても異質性の増加を導く可能性もある. TSD2 ではどの評価尺度が最も適切であるかは経験的なものであり、データに対する適合度に基づいて決定することは稀とあるが、TSD4 の著者らの経験によると、より高い I²統計量を示す評価尺度は、大きな不一致性を示すとのことである.

意思決定は、データに適合する一貫したモデルに基づいている必要があるものの、さまざまなエビデンスのソースを注意深く調べると、一部の推定値が裏付けされているものとそうでないものがあることが明らかになる場合がある。不一致性への対応方法を統計手法で決定することはできず、臨床疫学者による一貫したレビューが必要である。例えば用量が異なっている試験を特定したならば、研究者は次の一連の決定を行う必要があると TSD4 に記載されている:これらの要因は対象となる集団にどのように関係するのか?含まれるべきではない試験がないか?治療効果にバイアスがあるとみなし、調整すべきか、もしそうなら、調整の基となるデータはあるのか?もちろん、最終的な選択肢としては、明らかな不一致性や異質性の説明がつかないようであれば、偶然と考えることである。

参考文献

- [1] 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. available http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD4-Inconsistency.final_.15April2014.pdf
- [2] Bucher, H.C., Guyatt, G.H., Griffith, L.E., Walter, S.D. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. Journal of Clinical Epidemiology 1997; 50(6):683-691.
- [3] Chou, R., Fu, R., Hoyt Huffman, L., Korthuis, P.T. Initial highly-active antiretroviral therapy with a protease inhibitor versus a non-nucleoside reverse transcriptase inhibitor: discrepancies between direct and indirect meta-analyses. The Lancet 2006; 368(9546):1503-1515.
- [4] Cipriani, A., Furukawa, T.A., Salanti, G., Geddes, J.R., Higgins, J.P.T., Churchill, R. et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. The Lancet 2009; 373(9665):746-758.
- [5] Dias, S., Ades A.E., Welton N.J., et al. Network Meta-Analysis for Decision-Making. John Wiley & Sons, Inc. 2018.
- [6] Lumley, T. Network meta analysis for indirect treatment comparisons. Statistics inMedicine 2002; 21(16):2313-2324
- [7] Lu, G., Ades, A.E. Assessing evidence inconsistency in mixed treatment comparisons. Journal of the

- American Statistical Association 2006; 101(474):447-459.
- [8] Salanti, G., Higgins, J.P.T., Ades, A.E., Ioannidis, J. Evaluation of networks of randomized trials. Statistical Methods in Medical Research 2008; 17(3):279-301.
- [9] Dias, S., Welton, N.J., Caldwell, D.M., Ades, A.E. Checking consistency in mixed treatment comparison meta analysis. Statistics in Medicine 2010; 29(78):932-944.
- [10] Van Valkenhoef G, Dias S, Ades AE, Welton NJ. Automated generation of node-splitting models for assessment of inconsistency in network meta-analysis. *Res Synth Methods*. 2016;7(1):80-93.
- [11] Krahn U, Binder H, König J. A graphical tool for locating incosistency in network meta-analyses. BMC Med Res Methodol. 2013;13(1).
- [12] 奥山ことば,町田 光陽, 石渡 量太, 渥美 淳, 奥村 広之. 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. SAS ユーザー総会 2019.

5 ベースラインモデルにおけるエビデンスの統合(TSD5)

5.1 はじめに

本テクニカルサポートドキュメント (TSD) では、ベースライン群 (参照群又はプラセボ群) での自然病歴 (natural history:疾患の自然な進展;補足:寛解又は治癒する場合も含む) を表現したモデル (以下、ベースラインモデル) を構成する際に生じるエビデンスの統合問題を取扱っている.

TSD2[1]で、相対治療効果の統合に対して用いられた一般化線形モデリングのフレームワークを、「標準治療」又は「プラセボ」の対照治療の絶対的効果(補足:絶対的効果とは単群での効果を示している)のモデル化にも利用することが推奨されている。ベースラインモデルを構成するための情報源を正当化するべきであり、情報源には、系統的レビュー(システマティックレビュー)で特定された臨床試験、コホート研究、患者レジストリ、専門家の意見、又はこれらの組合せ等が考えられる。ベースライン群で観察されたバラツキをベースラインモデルに反映するため、固定効果又は変動効果モデルから得られた平均ではなく、予測分布を用いてベースラインモデルを構成することが推奨されている。

相対治療効果モデルが、ベースラインモデルの仮定により影響を受けないようにするため、相対治療効果モデルとは独立してベースラインモデルを構築することが好ましい.一方で、ベースラインモデル及び相対治療効果モデルの(互いに影響を及ぼす可能性のある)同時モデリングは、例えばエビデンスが非常に少ない場合、及び他の研究又は他の研究デザインが特定のベースラインモデルを信じる理由がある場合は、幾つかの利点がある.また、共変量の効果をモデル化する選択肢も利用可能とされている.

短期的な治療効果を超えた中長期的な効果を表すベースラインモデルでは、そのパラメータをエビデンス統合から得ることもある。最も単純で、最も容易に解釈可能なモデルは、短期的な治療効果のみを反映し、それ以降においては、独立した治療効果を考えないものである。しかしながら、エビデンスがそのモデル構造を支持しない可能性もある。その場合は、可能な限り異なる治療に対して、異なる値をとるすべてのパラメータは、無作為化試験から得られたデータに基づくべきである。短期間の治療効果と同様に、それ以降の群間比較に影響を与える非無作為化試験のエビデンスを用いる場合は、含めるかどうかに対する判断において細心の注意を払う必要があり、不確実性の増加及び潜在的な偏り(以下、バイアス)にも対処する必要がある。

さらに頑健(ロバスト)な結果を与え、かつモデルの内部妥当性を保証するため、可能であれば、臨 床試験データ、及び臨床試験データと観察研究データの組合せに基づく複数のベースラインアウトカム に対し、同時モデルを構成することが推奨されている.

状態推移モデル(マルコフモデル等)のための特別なエビデンス統合方法が利用可能で、異なる追跡時間での状態推移の情報、リスク(risks)(補足:薬剤疫学では、リスクとして「(発現)割合」を参照しており、生物統計で用いる用語とは異なっている)又は率(rates)として提示された結果、及び不完全な状態推移(つまり状態推移の途中で他の状態を経験する可能性がある場合)に関する情報の統合

に用いられる.

最終段階として、可能な限り、独立データを用いて、モデルの長期的な予測可能性を検証することが 重要となる。また検証(以下、バリデーション)データ又は較正(以下、キャリブレーション)データ の統合に基づき、モデルバリデーション及びキャリブレーションに対する複数パラメータによるエビデ ンス統合方法には幾つかの利点がある。

5.1.1 導入

ほとんどの費用効果分析(CEA)は、2つの別個のコンポーネントで構成されている。すなわち、比較対照群(comparator set)の標準治療のもとでの絶対的な自然病歴を表すベースラインモデルと、相対治療効果のモデルであり、これらは独立に推定することが望ましいとされている(後述)。前者は、臨床試験又は観察研究のコホートに関するエビデンスに基づいてもよいが、後者は一般に無作為化比較対照試験(RCT)データに基づく[2]。次に、新しい治療での自然な疾患推移(natural history)は、臨床試験データに基づいた相対治療効果の推定値と、ベースラインモデルを組合せることにより得られる。例えば、標準治療での望ましくない事象が発生する確率が0.25であり、標準治療と比較する所与の治療のオッズ比が0.8(治療群に有利)である場合、これらの量の不確実性を無視して(バラツキは無視して)、治療群での事象が発生する絶対的な確率は、以下の式よりp=0.21となる。

$$logit(p) = logit(0.25) + ln(0.8)$$

ここで、logit(x) = ln(x/(1-x))(詳細はTSD2[1]参照). 同様の方法を、対数相対リスク(対数尺度でのリスク比)又は対数ハザード(尺度)で線形であるモデルに対して利用することもできる[1].

通常,経済評価で,臨床試験データの役割 - 絶対的又は相対治療効果があるかどうか - ということは,短期的又は中期的なアウトカムに限定されている.

医療経済学者は、長期の治療効果をモデルにより構築することば、かなりの努力を費やしており[3]、そのモデルは、しばしば相対治療効果が特定の推移確率に基づき疾患の状態が推移していくと想定するマルコフ(状態推移)モデル構造となる。しかし、マルコフモデルのみならず、幅広いモデリング手法が利用可能である。より短期的アウトカムからの「マッピング」、又はマルコフステート(疾患状態)からQOL値(効用値)への「マッピング」は、さらに検討すべき事項とされているが、ここでは割愛する[3].

本TSDではベースラインモデルの構築におけるエビデンス統合に関する問題に焦点を当てる (NICEの手法ガイドの一般的原則[2]も参照).したがって、エビデンスの統合問題に影響する場合を除き、モデル構築の原則又はモデルタイプに関する推奨又はガイダンスを与えるものではない.経済モデルを通じて個々の患者を追跡するような患者レベルのシミュレーションモデルは、本TSDのスコープ外であり、エビデンス統合のみに焦点を当てる.より詳細は文献を参照[3][4].

まず比較研究(comparative study) (5.2節)でのアウトカムに関するベースラインモデルを検討する.

また,ベースラインモデルの推定において,どのようなエビデンスを利用すべきか,及びベースラインモデルと相対治療効果モデルとの潜在的な関連性を議論する.

5.3節では、ベースラインモデルに関連するエビデンス統合の問題で、臨床試験の期間を超える「下流 (downstream)」のデータ、相対治療効果以外のパラメータ情報をどのように得るかの問題を検討する. 推定が必要な典型的なパラメータは、基礎疾患から生じる合併症の発現率、入院期間、投薬期間、治療中断後の自然病歴、副作用の発現率、再発率、治療中または治療を中断/中止 (off treatment) での死亡率、サロゲートエンドポイントから臨床的エンドポイント (真のエンドポイント) への「マッピング」、又は疾患特異的な測定値からQOL尺度へのマッピングなど、多岐にわたる、幅広いモデルタイプ及びモデル構造を考慮することが可能である。マルコフの状態推移モデルを利用する場合、臨床試験データは、ある1つの状態から別の状態への推移(確率)のみにおいて利用可能であり、残りの推移(確率)に関する情報は、他のエビデンス (他の臨床試験など) から推定する必要がある。通常、これらのパラメータの推定に適切なデータを特定することは、選択されたエビデンスをどのように統合するかという技術的な問題よりも、意思決定においてより重要になるだろう。

しかし、2つの具体的な問題は、慎重に検討する必要がある. 理想的な場合では、すべての予測される治療間差は、無作為化比較対照試験(RCT)のデータに基づく. これは、通常RCTのデータに基づく短期的なアウトカムと同様に、その期間を超えた「下流」のアウトカムに対しても同様に適用される. 治療間比較に直接関係する非無作為化データの利用は、潜在的なバイアス[2]を慎重に考慮する必要が常にある(TSD3[5]も参照). 次に、「下流」アウトカムに関する情報が無作為化データ又は非無作為化データに基づいているかどうかに関して、観察された長期の相対的効果、及び短期的モデル並びにベースラインモデルにより予測された相対的効果との間に矛盾が生じる可能性がある. この分野はまだ十分に研究されていないものの、問題点を説明し、利用可能な方法を提案する.

最後に、5.4節では、モデルの「較正(以下、キャリブレーション)」と「検証(以下、バリデーション)」における問題点を簡単に説明する。「バリデーション」に関するほとんどの報告では、モデルから得られたアウトカムの予測値を、独立した情報源からのアウトカムが比較されている。「キャリブレーション」とは、外部データを考慮して、推定値を修正するプロセスである。第3の利用可能なエビデンス統合方法は、情報源とそれとは独立したデータの両方を同時に用いて推定するものである。5.4節では、バリデーションとキャリブレーションに対するエビデンス統合方法の性質を概説し、これらの方法への参考文献を与える。

5.2 臨床試験アウトカムに対するベースラインモデル

5.2.1 ベースラインアウトカムに関するエビデンスの情報源

ベースライン(又は参照群)の介入が定義されたら(TSD2[1]参照),系統的な研究調査(システマティックレビュー)及び試験の組入基準に関して,合理的なプロトコルを作成し[6][7][8],必要に応じ、代替オプション(補足:他の治療を参照群にする等)に関する潜在的な感度分析を規定すべきである(補足:相対効果はRCTを参照することで相対効果を安定して推定可能であるが、ベースラインモデルでは

外的妥当性が重視される). ベースライン(又は参照群)での反応は、対象となる集団に対して可能な限り固有でなければならないため[2][3]、最近の試験、関連するコホート研究、レジストリ研究[9]、又は場合によっては専門家の意見[8]からのエビデンスを利用する方が、より合理的かもしれない. ベースラインのアウトカムに関する情報源を特定するための共通の方法は、相対的効果が得られた臨床試験と同じ試験を利用する事であり、その場合、ベースライン治療を行った投与群の情報のみが利用される. これは利用可能な方法ではあるが、各々のケースにおいて正当化される必要がある. 研究者は、相対的効果を得るのに用いられた全ての臨床試験が、対象患者集団及び現在の環境の下で得られる絶対的応答を等しく代表していると考えられるかどうか、特に幾つかの臨床試験は何年も前に実施されたかどうか、又は非常に限定的な組入基準を採用していたかどうかを考慮すべきである. 異なるタイプの無作為化試験及び非無作為化試験のエビデンスを組合せることも可能である.

意思決定モデルを構築するために利用されたエビデンスの情報源が何であれ、透明性があるものでなければならず、外部の検閲を可能にするために十分な詳細が報告されるべきである[2][6][8][10].

5.2.2 ベースライン反応に関する要約データの統合

ベースラインと治療効果に対する別々のモデル

TSD2[1]では相対的効果を統合するため、一般化線形モデルの枠組みを導入した.これは次のように表せる.

$$g(\gamma) = \theta_{ik} = \mu_i + \delta_{i,1k} I_{\{k \neq 1\}}$$

$$I_{\{u\}} = \begin{cases} 1 & u i i 真 \\ 0 & その他 \end{cases}$$

g()はリンク関数(例えば、logitリンク関数)、 θ_{ik} は線形予測子、 μ_i は臨床試験iでの試験固有のベースライン効果(例えば、対数オッズ)、及び群1に対する群kの治療効果 $\delta_{i,1k}$ (例えば、対数オッズ比)で構成される。相対治療効果のモデルに対するベイズ流の枠組みでは、 μ_i は交換可能性を仮定しない無関係な(unrelated)漠然事前分布が与えられる。一方、ベースライン効果をモデル化するためには、以下の定式化を採用する。

$$g(\gamma) = \theta_{1k} = \mu_i$$

$$\mu_i \sim N(m, \sigma_m^2) \tag{1}$$

ここで、試験固有のベースラインは、相対効かとは異なり、共通の平均及び分散を有するベースライン効果の分布から導き出される(交換可能性を仮定する). ベイズ流のモデルでは、平均及び分散に対し漠然事前分布を与える. 例えば、 $m \sim N(0,100^2)$ 、 $\sigma_m \sim Uniform(0,5)$ 又は $1/\sigma_m^2 \sim Gamma(10^{-3},10^{-3})$ である.

すなわち, 本TSD5での提案とは, 相対的効果を推定したモデルとは別のモデルを実行し, ベースライ

ンのデータを要約することである.1つのオプションは,ベースラインモデルの情報が相対治療効果モ デルに伝播しないことを保証しながら、相対治療効果モデルと同時に、ベースラインを推定するコード を実行することである.これはcut機能を用いてWinBUGSで実行可能である[11].本方法の利点は,両方 のモデルが1つのプログラムファイルに含まれ、ベースラインモデルに追加された新しいデータが、相 対的効果モデルから得られる絶対的効果 (ベースライン効果) を自動的に更新することを保証しつつ, 同時に実行可能であることである. また, これはベースライン効果の事後分布の標本が, 直接利用でき ることも保証している. より単純なもう1つのオプションは、別々のモデルを実行し、ベースライン効果 の事後分布の正規性を仮定し、その事後分布から適切な要約統計量[すなわち、平均と不確実性(=バ ラツキ)]を算出し、それらを相対的な効果を推定するプログラムコード(相対治療効果のモデル)に 代入することである.これは当然、ベースライン効果の事後分布の正規近似に依存する.これは常に チェックされるべきだが、通常、NICEの経験上では成立している。あるいは、ベースライン効果の事後 分布からの標本を、別で実行する相対治療効果モデルに投入することも可能である. TSD2[1]の付録の 各事例では、様々なモデルとアウトカムのタイプ(連続値、2値など)に対するWinBUGSコードが与え られており、そのコードに外部データから別途推定した結果を「プラグイン」して(差し込んで)、絶 対的な(ベースライン群の)応答率又は反応確率の予測を行っている.説明目的のため、外部データの 解析では、変量効果モデルに基づき、「新しい」試験でのベースライン効果の予測分布を生成している. これは、相対治療効果モデルでのベースライン群の重み付けされていない平均値を単純に計算する方法、 つまりどのような状況でも推奨されない方法,とは異なることに注意する.

ベースラインモデルと相対治療効果モデルを別々に設定する理由は、ベースラインモデルに対する仮定が、相対治療効果に影響を及ぼすことを回避するためであり、また2つのモデルが異なる情報源に基づくことが多いためである。しかし、特定のベースラインモデルを信じる強い理由がある場合、ベースラインモデルと相対治療効果モデルの同時モデリングが考慮されるべきである(5.2.3節参照). アウトカム間の相関関係、及びベースラインと治療効果との交互作用に関連する問題は、TSD2[1]及びTSD3[5]でそれぞれ議論されている.

解析を実施するには2つの方法がある.最も単純な方法は、mの事後平均及びその事後標準偏差を利用することである.しかし、これがデータから観察されるバラツキを過小評価することが議論になり得る.すなわち、もしベースライン群のデータをよりもっと集めることができれば、平均の推定値はよりもっと正確になるが、そのバラツキは変わらないままであろう.よって、その代替法としては、以下の新しいベースラインの予測分布を利用する方法である.

$$\mu_{new} \sim N(m, \sigma_m^2) \tag{2}$$

ここで、 $m \ge \sigma_m^2$ は事後分布からサンプリングされ算出される。新しいベースラインの予測分布は、データで観測されたバラツキと同様に、新しい観測値の不確実性も含んでいる。しかし、予測分布により伝達される不確実性が、ベースラインの真の不確実性を確実に反映していることが重要である。したがって、ベースラインモデルにどの試験を用いるべきか、及び含まれている試験のベースライン効果と「新しい」ベースライン(式(1)及び(2))との間の交換可能性の仮定が成立しているかを、慎重に評価する必要がある。

以下に,実行プログラム(プログラム1)を示す.禁煙データセット[12](TSD4[13]の4.2.1節参照)から19個の「介入なし」群を含み,式(1)を実行するコードが含まれている.なお,WinBUGSプログラムはTSD発行以降に修正されている場合があることから,以下のWebサイトから最新プログラムをダウンロードすることを強く推奨する.

ダウンロード: http://nicedsu.org.uk/technical-support-documents/evidence-synthesis-tsd-series/

プログラム1. 禁煙の事例: Binomial likelihood, 予測分布を有するベースライン変量効果モデル

```
# Binomial likelihood, logitリンク
# 変量効果によるベースラインモデル
                                      # *** PROGRAM STARTS
model{
                                # 試験のループ処理
for (i in 1:ns){
r[i] \sim dbin(p[i],n[i])
                                      # 尤度定義
logit(p[i]) \leftarrow mu[i]
                                      # 対数オッズ
mu[i] ~ dnorm(m,tau.m)
                                      # 変量効果モデル
                                      # 予測分布(対数オッズ)
mu.new ~ dnorm(m,tau.m)
                                # 平均に対する漠然事前分布
m \sim dnorm(0,.0001)
<u>var.m <- 1/tau.m</u>
                                      # 試験間分散
                                      # 試験間精度=(1/試験間分散)
tau.m <- pow(sd.m,-2)
                                       # 試験間標準偏差の漠然事前分布
sd.m \sim dunif(0,5)
```

ベースライン予測平均 (mu_new の事後平均値) 及び不確実性の推定値 (mu_new の事後分散) を相対的治療効果のモデルに入力し、絶対的な確率が計算可能である (TSD 2[1]の付録に詳述).

ベースライン変量効果の分散に対しては、その他の事前分布も利用可能である(事前分布はTSD2[1] の6.2節参照). 例えば、精度パラメータに対して、漠然ガンマ事前分布に置き換え可能であり、分散に対する漠然逆ガンマ事前分布に対応する.

```
tau.m ~ dgamma(0.001,0.001)
sd.m <- sqrt(var.m)
```

以下のコードを右中括弧(})の前に追加することで、ベースラインの対数オッズの平均の事後分布 (R) 又は予測分布 (R.new) に基づいたベースラインの反応確率を推定可能である.

```
logit(R) <- m # 事後確率
logit(R.new) <- mu.new # 予測確率
```

データ構造には2つの要素があり、研究数 ns を指定したリスト、及びベクトル形式のデータそのものであり、r[] 及び n[] はそれぞれベースライン群での分子と分母である。プログラム実行にあたり、両方のデータをWinBUGSにロードしておく必要がある。

データと連鎖ごとの初期値

```
# Data (Smoking Cessation: baseline arms only)
list(ns=19) # ns=number of studies
r[]
   n[]
        # Study ID
   140
        # 1
9
75
   731
        #3
   106
        #4
2
(中略)
        # 19
78
  584
69
   1177
        # 20
END
# Initial values
#chain 1
list(mu=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0), sd.m=1, m=0)
#chain 3
```

表 5-1の第1列は禁煙の例において、ベースライン効果、及び相対治療効果に対し、別々の変量効果 (RE) モデルを用いて得られた結果を示す。これらのすべての結果は、収束性を確実にするために20,000 回のバーンイン (Burn-in) サンプルを破棄した後、3つの独立した連鎖(チェーン)から50,000回の反復に基づいたものである。平均値の事後分布を用いると、禁煙率の平均ベースラインが0.07、及びその信用区間 (0.05,0.09) が得られる。対照的に予測分布を用いる場合、平均値はほぼ同じ (0.08) であるが、正規的に分布する変量効果の仮定の下では、観測データの変動範囲をよりよく反映して(すなわち、平均の分布ではなく、データの分布そのものを反映して)、より広い信用区間 (0.02,0.20) となっている (表 5-1) 。

表 5-1 ベースラインと治療効果を別々にモデリング又は同時にモデリングした変量効果メタアナリシスを実行. 「No Contact (介入なし)」の禁煙率の(予測対数オッズの)平均値及び予測対数オッズ(m 及び μ_{new})の事後平均,標準偏差 (sd) 及び 95%信用区間 (CrI),各群の対数オッズの事後分布及び予測分布から算出した (絶対)禁煙率の事後平均,標準偏差 (sd) 及び 95%信用区間 (CrI) ,及び各群の「No Contact (介入なし)」に対する対数オッズ比の事後平均,標準偏差 (sd) 及び 95%信用区間 (CrI) (対数オッズ比>0 ;治療群に有利な方向). ベースラインの試験間の異質性 (σ_m) 及び治療効果の試験間の異質性 (σ) に対する事後中央値,標準偏差 (sd) 及び 95%信用区間 (CrI) .

モデル適合度:残差逸脱度(resdev)の事後平	·均、パラメータの有効数(pD)及び DIC.
--------------------------	-------------------------

	Separate Models		Si	multaneous n	nodelling	
	mean	sd	95% CrI	mean	sd	95% CrI
	/median			/median		
			Baseline mo	odel parameter	rs	
m	-2.59	0.16	(-2.94, -2.30)	-2.49	0.13	(-2.75, -2.25)
σ_m	0.54	0.16	(0.32, 0.93)	0.45	0.11	(0.29, 0.71)
μ_{new}	-2.59	0.60	(-3.82, -1.41)	-2.49	0.49	(-3.48, -1.52)
		Absolut	e probabilities of r	esponse based	on the poste	<u>rior</u>
			<u>distribution</u> of the	e baseline prol	bability	
No contact	0.07	0.01	(0.05, 0.09)	0.08	0.01	(0.06, 0.10)
Self-help	0.12	0.05	(0.05, 0.23)	0.13	0.04	(0.07, 0.21)
Individual Counselling	0.15	0.04	(0.09, 0.24)	0.15	0.03	(0.11, 0.21)
Group Counselling	0.19	0.07	(0.08, 0.37)	0.20	0.05	(0.11, 0.31)
	Absolute probabilities of response based on the <u>predictive</u>		<u>ctive</u>			
			distribution of the	e baseline prol	bability	
No contact	0.08	0.05	(0.02, 0.20)	0.08	0.04	(0.03, 0.18)
Self-help	0.13	0.08	(0.03, 0.34)	0.14	0.07	(0.04, 0.30)
Individual Counselling	0.17	0.09	(0.05, 0.39)	0.16	0.07	(0.06, 0.33)
Group Counselling	0.21	0.12	(0.05, 0.50)	0.21	0.09	(0.07, 0.43)
		Relat	ive treatment effec	ts compared to	o 'No contac	t'
Self-help	0.49	0.40	(-0.29, 1.31)	0.53	0.33	(-0.11, 1.18)
Individual Counselling	0.84	0.24	(0.39, 1.34)	0.78	0.19	(0.41, 1.17)
Group Counselling	1.10	0.44	(0.26, 2.01)	1.05	0.34	(0.39, 1.72)
σ	0.82	0.19	(0.55, 1.27)	0.71	0.13	(0.51, 1.02)
resdev*	54.1			47.4		
pD	45.0			40.1		
DIC	99.1			87.5		

^{*} データポイント50と比較

[出典: NICE DSU TSD5 (http://nicedsu.org.uk/technical-support-documents/)]

事後分布又は予測分布どちらを選択するかは、治療間差にはほとんど影響を与えないが、予測分布はベースラインモデルのより大きな不確実性(バラツキ)を反映していることに留意する。事後分布と予測分布の不確実性の両方を用いて計算された4つの治療の禁煙率を図 5-1に示す。図 5-1は、2つの方法が治療効果の不確実性の推定に及ぼす影響を示しており、予測分布を用いた場合、信用区間がより広がっている。将来の予測データに対する不確実性の真のレベルを表している。

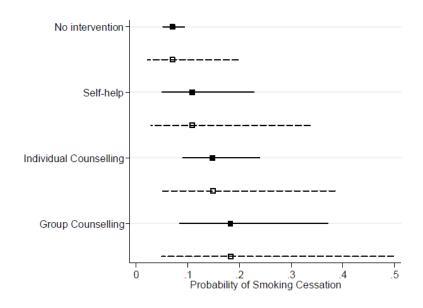


図 5-1 ベースライン及び治療効果に対し別々の変量効果モデルを構築した場合. 各治療群での(絶対) 禁煙率の平均 対数オッズの「事後分布」に基づく95%信用区間(■,実線)及び「予測分布」に基づく95%信用区間(□,破線).

[出典: NICE DSU TSD5 (http://nicedsu.org.uk/technical-support-documents/technical-support-documents/)]

ベースライン効果と相対治療効果の同時モデリング

絶対的な治療効果,及び相対治療効果を分割することは,人工的に見えるかもしれないが,それでも相対治療効果がベースラインに対する仮定の影響を受けないため,推奨される方法である.しかし,ベースライン効果と治療効果を一緒にモデリングする理由があるかもしれない.1つの理由は,群とアウトカム(2値)の分割表のセルが零(ゼロ)の場合,モデルの安定性を向上させるためである(TSD2[1]6.3節参照).もう1つは,他の研究に基づいて,例えば,クラスター内での無作為化[14][15],又は多施設試験の結果をモデリングする場合など,ベースラインに特定のモデルを信じる強い理由がある場合(例えば、臨床的に妥当なモデル)である(このような場合,通常,ベースラインに対するモデルは仮定しない.TSD2[1]参照).このような分析を実施するには,TSD2[1]に示された標準的なメタアナリシスのコードにおけるμに対し,交換可能性を仮定しない「無関係な」事前分布ではなく,交換可能性を仮定する平均と分散のある「変量効果」の事前分布で置き換え、ベースライン効果の平均及び各試験間の分散の事前分布を与えることだけが必要となる.全ての試験がベースライン (参照)治療を含むわけではないネットワークメタアナリシスでは,モデル化されたμが常にベースライン治療,すなわち治療1となることを保証する必要がある.ベースラインと相対治療効果を同時にモデリングする場合は,相対治療効果の推定値に大きな影響を与えるため,常に正当化される必要があることに留意する.可能であれば,相対治療効果への影響を検討する感度分析を実施すべきである.

以下にベースラインと相対治療効果の同時モデリングのWinBUGSコード(プログラム2)を示す.ここで再度,「新しい」ベースライン(式(2))の予測分布を使用することを推奨する(すなわち,事後分布ではなく,予測分布を使用すべき).

ベースライン効果のモデルを含めることは、相対治療効果の事後分布に大きな影響を与える可能性がある。したがって、前述したように、非常に特殊な状況下でない限り、本モデルは推奨しないとされている。本モデルを使用する場合は、詳細に正当化されるべきであるとされている。

プログラム2. 禁煙の事例: Binomial likelihood, ベースライン効果と治療効果を, 同時に予測分布により変量効果でモデル化

```
# Binomial likelihood, logit link
#3群以上の臨床試験に対するベースラインと治療効果の同時モデル
                                      # *** PROGRAM STARTS
model{
for(i in 1:ns){
                               # 試験のループ処理
                               #3群以上の臨床試験の補正(対照群は0)
w[i,1] < 0
delta[i,1] <- 0
                               # 対照群の治療効果は0
                                      # 試験ベースライン(参照治療1)
mu[i] ~ dnorm(m,tau.m)
for (k in 1:na[i]) {
                                      # 治療群のループ処理
   r[i,k] \sim dbin(p[i,k],n[i,k])
                                      # 二項尤度
  logit(p[i,k]) <- mu[i] + delta[i,k]
                                     # 線形予測に対するモデル
   rhat[i,k] \leftarrow p[i,k] * n[i,k]
                                      # 分子の期待値
   dev.NA[i,k] \leftarrow 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
                                            # NAを含む逸脱度
     + (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
   dev[i,k] <- dev.NA[i,k]*(1-equals(n[i,1],1)) # NAに対する逸脱度の補正
resdev[i] <- sum(dev[i,1:na[i]])
                                      # 試験ごとの残差逸脱度の合計
                                     # 治療群のループ処理
for (k in 2:na[i]) {
   delta[i,k] ~ dnorm(md[i,k],taud[i,k])
                                      # 試験特有の対数オッズ比(LOR)の分布
   md[i,k] \leftarrow d[t[i,k]] - d[t[i,1]] + sw[i,k]
                                    #LOR事後分布の平均(3群以上の試験の補正)
   taud[i,k] <- tau *2*(k-1)/k
                                     #LOR事後分布の精度(3群以上の試験の補正)
  w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]]) #3群以上の臨床試験の補正
   sw[i,k] <- sum(w[i,1:k-1])/(k-1)
                                      #3群以上の臨床試験の累積補正
}
}
totresdev <- sum(resdev[])
                                     # 残差逸脱度の総合計
                                     # 参照治療1の治療効果は0
for (k \text{ in } 2:nt) \{ d[k] \sim dnorm(0,.0001) \}
                                    # 治療効果の漠然事前分布
sd \sim dunif(0,5)
                                     # 試験間標準偏差の漠然事前分布
tau \leftarrow pow(sd,-2)
                                      # 試験間精度=(1/試験間分散)
mu.new ~ dnorm(m,tau.m)
                                      # ベースライン(対数オッズ)の予測分布
m ~ dnorm(0,.0001)
                               # 平均の漠然事前分布(ベースラインモデル)
                                      # 試験間分散(ベースラインモデル)
<u>var.m <- 1/tau.m</u>
                                       # 試験間精度=(1/試験間分散)
tau.m <- pow(sd.m,-2)
                                       # 試験間標準偏差の漠然事前分布 (ベースラインモデル)
sd.m \sim dunif(0,5)
                                # *** PROGRAM ENDS
}
```

右中括弧(})の前に、以下のプログラムコードを追加することで、治療1(ベースライン/参照治療)のベースラインでの対数オッズ平均の事後分布、又は予測分布に基づいた各治療群の絶対効果の推定値を算出できる。

```
# ベースラインモデルの事後分布に基づく治療効果 T[k] の推定値,
# 及びベースラインモデルの予測分布に基づく治療効果 T.new[k] の推定値を算出

for (k in 1:nt) {
    logit(T[k]) <- m + d[k]
    logit(T.new[k]) <- mu.new + d[k]
    }
```

データと連鎖ごとの初期値

```
# Data (Smoking Cessation)
# nt=no. treatments, ns=no. studies
list(nt=4,ns=24)
r[,1] n[,1]
              n[,2]
                    r[,3]
                         n[,3]
                               r[,4]
                                    n[,4] t[,1]
                                               t[,2]
                                                     t[,3]
                                                          t[,4]
                                                                na[]
                                                                     #ID
         r[,2]
   140
              140
                                                                      # 1
         23
                    10
                         138
                               NA
                                    NA
                                               3
                                                     4
                                                           NA
                                                                3
9
                                          1
NA 1
         11
              78
                    12
                         85
                               29
                                    170
                                          1
                                               2
                                                     3
                                                                4
                                                                      # 2
   731
         363
             714
75
                   NA
                         NA
                               NA
                                    NA
                                          1
                                               3
                                                     NA
                                                           NA
                                                                      #3
(中略)
NA 1
                                                                     # 23
         12
              76
                    20
                         74
                               NA
                                    NA
                                          1
                                               3
                                                     4
                                                           NA
                                                                3
NA
         9
              55
                    3
                         26
                               NA
                                    NA
                                          1
                                               3
                                                     4
                                                           NA
                                                                3
                                                                      # 24
END
# Initial values
list(sd=3, m=.5, sd.m=.5, d=c(NA,-2,5,-5), mu.new=-1, mu=c(-1,5,-3,1,-1,5,1,2,3,2,1,5,2,1,-5,1,2,-5,-3,1,5,2,1,-5))
```

WinBUGSプログラムは、TSD発行以降に修正されている場合があることから、以下のWebサイトから 最新プログラムをダウンロードすることを強く推奨する.

ダウンロード: http://nicedsu.org.uk/technical-support-documents/evidence-synthesis-tsd-series/

試験数 ns , 治療数 nt , r[,1] 及び n[,1] は最初の治療(その試験での対照群;補足:ベースライン 群でも参照薬群でもない場合もあることに注意)の分子及び分母である. r[,2] 及び n[,2] は2番目の治療群の分子及び分母であり, r[,3] 及び n[,3] は3番目の治療の分子及び分母である. t[,1], t[,2] 及び t[,3] は各試験で比較されている治療法(番号)であり, na[] は当該試験での治療群数を示す.

本事例では「連絡しない (No Contact)」がベースライン/参照治療として選択された. 幾つかの試験で、ベースライン治療1が含まれていなかったことに注意する必要がある (試験2, $21\sim24$). モデルが正しくベースラインパラメータ mu を推定できるように、r[,1]=NA, n[,1]=1 及び群数について、これらの試験に治療1を加えることで修正している.

表 5-1の第2列は、ベースライン効果と治療効果の同時モデルからの4つの治療群での禁煙の事後確率と予測確率である。ベースライン効果と治療効果を同時に推定するモデルは、試験間の異質性の推定値に大きな影響を与え(別々のモデルにおける事後中央値は0.82、同時モデルの事後中央値は0.71)、結果として平均治療効果の不確実性に大きな影響を与える。さらに、予測分布に基づく絶対的な治療効果の

不確実性も小さくなる.

観察されたベースライン効果の異質性 σ_m は、ベースライン効果と治療効果を別々に解析した場合より、同時にモデル化した場合の方が小さくなる。 表 5-1の μ_{new} の標準偏差が与えられたもとで、ベースライン効果の予測分布のバラツキに影響する.

5.2.3 共変量を伴うベースラインモデル

要約データの利用

共変量は、線形予測子に変数を追加することで、ベースラインモデルに含めることができる.共変量 Cは、連続変数の共変量、又はダミー変数の共変量のいずれかの場合、試験iの群kに対して以下を考える ことができる.

$$\theta_{ik} = \mu_i + \beta C_i + \delta_{ik} I_{\{k \neq 1\}}$$

推定値 μ_i と同様に、共変量効果 β の推定値は、臨床試験データ又は外部データから得ることができる. Govanら[16]は、共変量が要約バイアス(aggregation bias)[17]を減らす目的で、臨床試験の要約データから推定する例を与えた、要約バイアスとは、強い共変量が存在する場合、例えその共変量の分布が投与群でバランスが取れていても、且つその共変量が相対治療効果に対する効果修飾因子でない場合であっても、相対治療効果の推定値をヌル(無)方向(効果がない方向)に偏らせる現象である. Govanら[16]は、共変量の欠損データを取扱う方法も示している(詳細はTSD3[5]参照).

個別患者データ(IPD)によるベースラインモデルにおけるリスク方程式

年齢、性別、治療開始時の疾患の重症度などの共変量によって、ベースラインの進行の違いを説明できるベースラインモデルを構築する上で、よりもっと信頼できる方法は、個々の患者データを利用するものである。(回帰)係数がより精確に推定され、生態学的バイアス(ecological bias)のリスクをより軽減させることができるため、要約データを用いる方法よりも(患者の個別データを用いた方が)優れると考えられる。その結果は、大規模な試験データベース研究、レジストリ研究、又はコホート研究からの複数の回帰に基づいた「リスク方程式(risk equations)」として表現される。そして、あたかもそれらは別の危険因子であるかのように、臨床試験データに基づく相対的な治療効果を、リスクの方程式に単純に加えることにより、各治療群の自然病歴が生成される。ここで、費用対効果の分析者が直面する主たる難点は、情報源の選択及び対象患者集団の適切性を正当化することである。これら代替試験の患者集団での様々な特徴、及び意思決定に関わる対象集団との関係を検討する分析が提示されなければならない。必要に応じて、これらパラメータを得るのに利用された情報源の選択に関して、結果の感度を示す目的で、感度分析を提示する必要がある。

5.3 ベースラインモデルの他の部分における統合の問題

エビデンスの情報源の選択、及び直近の短期試験のアウトカムを超えた自然病歴に関する統計モデル

は、本文書のスコープ外である.しかし、これはエビデンス統合の問題、モデルの内部一致性、及びエビデンスとの整合性に関わるため、より長期間でのアウトカムにおける治療群間差、又は示唆された治療群間差の原因に関して、幾つかコメントする.

5.3.1 ベースラインモデルのパラメータ及び相対的な治療効果に対する情報源

一般的に、自然病歴の各パラメータに利用されるエビデンスの情報源は、プロトコル・ドリブン・レビュー(プロトコルを作成した上でレビューを行う)により、決定されるべきである[2][6][8]. 以前実施されたCEA(費用効果分析)は、自然病歴の重要な情報源である(先行するCEAを参照するべし).

共通するモデリング方法は、より短期間の試験アウトカムに関する条件付きで、「下流 (downstream) の」モデルでは治療間差がないと仮定するものである. 短期間の治療間差を与えたもとで、すべての治 療群に適用可能な単一マッピングによって、より長期間の治療間差が得られることから、これを「単一 マッピング仮説 (single mapping hypothesis) 」と呼ぶ、例えば、インフルエンザ治療薬の様々な抗ウイ ルス薬の費用対効果を分析するモデルでは、ベースケースの解析では、抗ウイルス薬の使用が短期間の アウトカムにのみに影響し、より長期間の合併症や入院率への追加効果(影響)はないと仮定している [18]. この性質は魅力的であるが、強い仮定を置いている. 代替の比較薬が単一のクラスと見なせれば、 この仮定は自然であるが、そうでない場合は妥当性が低い可能性がある.このような仮定は、臨床的及 び生理学的に正当化されなければならず、「マッピングされた」各アウトカム、例えば、入院期間、治 療期間、合併症の発現率、死亡率及び他のすべての下流(downstream)のアウトカム等に対し、仮定は レビューされ、検討され、及び解釈されなければならない. そのレビューには、サロゲートアウトカム (代替アウトカム)の妥当性に関する経験的及び統計的な文献も含むべきであり,特に,そのエビデン スが、より短期的な「サロゲート(代替)」アウトカムで検討した治療効果が、すべての治療群で長期 的な有益に変換されるという考え方を支持するか検討すべきである。このレビューは、「単一マッピン グ」 仮説が成立する治療群の範囲が広ければ広いほど、より頑健(ロバスト)である可能性が高いため、 考慮する製品クラスを超えて、拡大できる可能性がある。しかし最終的には、サロゲート(代替)アウ トカムと臨床アウトカム(真のアウトカム)との関係性は、意思決定内の治療のサブセットに対しての み適切と判断される可能性がある.

医療技術評価 (HTA) における「サロゲートエンドポイント」利用の議論は、臨床的及び統計的文献で「サロゲート」として古典的に理解されているアウトカムとは大きく異なる[19]. モデル化する治療間差は短期的なエビデンスに基づくことが許容されており、HTAの文献では、短期間のアウトカムの結果を長期のアウトカムへ「マッピング」することを頻繁に行っている。モデルの下流のアウトカムにおけるすべての治療間差がほぼ短期間の臨床試験アウトカムの治療間差によるものであるという仮定が、エビデンスによって支持されない場合、第1の選択肢は、より長期のアウトカムが得られ、利用可能な無作為化されたエビデンスを利用することである。これは必然的に、それぞれの治療に対して異なる「マッピング」を行うことを意味する。第2の選択肢であり、最優先の選択肢は、無作為化されていないエビデンスの利用である。しかし、短期間のアウトカムと同様に、非無作為化データのあらゆる利用は、モデル内の治療効果の差異に直接的に影響を与えるものであり、注意深く正当化され、不確実性が増加し、バイアスの可能性を認識し、対処することが不可欠となる[2].

5.3.2 ベースラインモデルを構成するための複数アウトカムの同時統合

ベースラインモデルは、通常、一連の「状態」又はサブプロセスで構成され、生涯コスト、QOL及び 生存年に影響を与える一連のパラメータを包含している. これらのパラメータを利用可能なすべての データから同時に推定することが望ましい.そうすることで,より多くの情報を取り込み,モデル予測 とエビデンスとの一致性に関して、より妥当性の検討が実行できるためである. 複数アウトカムの首尾 一貫したモデリングの最も単純な例は,本書では割愛したが,TSD2[1]に記載されている競合リスク及 び順序プロビット解析により与えられる. 例えば、ベースライン効果及び治療効果に対し順序プロビッ トモデルを利用することは、患者がACR(リュウマチの評価基準)などのカテゴリ尺度で、異なるレベ ルの応答を示す確率に対し、一貫した予測値を保証する. 対照的に、ACR 20、ACR 50、及びACR 70(そ れぞれARC基準20%,50%及び70%の改善あり)の応答を別々に分析すると、結局、予測不可能なモデル になってしまう可能性がある. 例えば、ACRの20%の改善を経験する患者数よりも、ACRの50%の改善 を経験する患者数の方がより多くなるようなモデルになってしまう可能性である. しかし, 高度なモデ リング技術の利用は、費用効果分析に大きな影響を与えないかもしれず、各自然病歴のパラメータが、 データから独立して決められる通常の方法は、より一般的に採用されている. 相対的な治療効果を得る ための複数の試験アウトカムを同時モデリングすることは、HTAにおいてますます見られるように、特 定の利点がある.異なるアウトカムの「一貫性のある」見解を反映し、それらアウトカム間の相関関係 を正しく捉えることと同様に、これらの方法は、異なる臨床試験で異なるアウトカムが報告されるとい うよく発生する問題に対処できる.治療間比較に対するベースケースとして単一のアウトカムを選択す るオプションは、廃棄される情報の割合が高くなる可能性がある。モデルが異なるアウトカム間の関係 性を考慮し、それゆえ治療効果に関するすべてのエビデンスを取り込むことができるならば、より頑健 (ロバスト) な結果が得られるため、望ましいだろう. 複数のアウトカムに対する治療効果のモデルの 例には、複数のフォローアップ回数での治療効果[20][21]、及び連続値のアウトカムに対する多変量モデ ルが含まれる[22][23][24]. 2つの試験でのアウトカム及びアウトカム(しかし、観察データに基づく)に 関連するパラメータ[25][26]を統合することも可能である. より複雑な例は, インフルエンザ治療薬の解 析[18][27]であり、「発熱の終了までの時間」と「症状の終了までの時間」の関係性のモデル、又は進行 性乳がんでの腫瘍反応(率),無増悪期間,及び全生存期間(OS)のアウトカムの統合などである[28][29]. しかし、モデル構造はさまざまな疾病によって異なり、同じ種類の構造であっても、モデルを構築する ために利用できるエビデンスの構造はかなり異なり得る。これらの理由から、幾つかのアウトカムを含 む単一のモデルは、その仮定が明確であり、臨床専門家において合意形成された見解を反映している限 り、費用対効果モデリングに対するより頑健(ロバスト)な基盤を与える可能性が高いことに留意する. それ以外に関しては、一般的な推奨を与えることは困難である.

状態推移モデル(state transition models)の統合

前述したように、全ての適切な情報源から入手可能な情報を取り込める方法の利用には利点があるが、他のベースラインモデルと同様に、状態推移モデルのパラメータは、それぞれ異なる情報源から独立して推定したり、同時にモデル化したりする.しかしながら、状態推移モデルのパラメータの統合は、情報が利用可能となる形態は多種多様であるため、例えば、以下のような幾つかの特別な考慮点がある.

- 1. 試験 j のデータは、モデルを構築するモデラーがサイクル時間 T_0 のモデルにおいて、これらのデータを利用したい一方、時間間隔 T_j での状態推移確率として報告される。標準的な調整方法[30]は、2 つの状態モデルに対してのみ妥当であることに留意することが重要である.
- 2. 情報はリスク (割合) 又は率で入手可能である.
- 3. 情報は「ハザード比」として利用可能かもしれないが、「相対リスク」はサイクル時間に依存する ため、多状態(マルチステート)モデルでの「相対リスク」に容易に変換することができない(そ の逆もしかり).
- 4. 状態Aから状態Bへの状態推移に関する情報が利用可能であるが、各個人がその間(AからBの間) に他の状態を経験する可能性があるような状況の場合、これは不完全に観察されたマルコフ過程 と呼ばれることもある.

様々な方法で、異なる期間を通じ、モデルの様々な状態間で報告される推移に関する幅広い情報を統合する方法がある[31]. さらに、これらの方法は、前述したように、自然病歴及び相対的な治療効果パラメータを同時にモデル化するために利用できる[32]. そのような方法は、以下で述べるキャリブレーション(調整)するための統合方法の実例も与える。現在まで、これらの方法は、すべての推移時間が指数的に分布している場合に限定されていた。どのような条件の下で、いかにしてその方法を他の分布に拡張可能かはまだ研究中である。

5.4 複数パラメータ統合によるモデルのバリデーションとキャリブレーション

ベースラインモデルは、可能な限り独立したデータに対してバリデーションされるべきである. 例え ば、新しい抗がん剤と標準的な対照薬とを比較する費用対効果評価では、標準的な対照薬群で予測され た生存率は、恐らく年齢又は他の共変量の適切な調整後の公表された生存率と比較できる.他の状況で は、人口規模に関する統計とともに、発生率又は罹患率に関する推定値の初期値を与えた条件で、ベー スラインモデルは、特定の後遺症、合併症、又は死亡を伴い病院に入院した患者の絶対数に関する予測 値を提供する.繰り返しになるが、バリデーションの一形態を提供するため、これらの予測を独立した データと照合できるだろう. より洗練された方法は、この外部データを利用してベースラインモデルを 「キャリブレーション」(較正)するものである.これは、モデルが独立したキャリブレーション(較 正) データを正確に予測できるように、モデル内の「進行速度の」パラメータを変更することである. 特にベイズ流の枠組みのキャリブレーションは、エビデンス統合の一形態ともとらえられる[33].この 場合、データをキャリブレーション(較正)することは、モデルパラメータの複雑な関数の推定値を提 供するものとして見なす、原則として、必要とされることのすべては、データをキャリブレーションす ることで推定するモデルパラメータの関数を研究者が指定することであり、さらに追加データの尤度の 項をモデルに追加することであり、この方法はキャリブレーションの極めて簡単な形態を提供するもの である. そして, 基本パラメータの推定値を得るため, モデルを介して(特定の期間以降の)「下流へ」 情報が伝播する. Vanniら[34]によってレビューされているキャリブレーションの標準的な方法より、本 方法には多くの利点がある.

- 1. サンプリング誤差を考慮して、キャリブレーションしたデータに適切な重みを与える.
- 2. キャリブレーションしたデータに「適合」するまで、モデルパラメータの「微調整」する方法、つまり、データの不確実性を捕捉できない方法を回避することができる.
- 3. 研究者が幾つかのベースラインモデルのパラメータのうちのどれを変更すべきかを, (恣意的に) 決定することを回避する(下記参照).
- 4. データバリデーションがモデルの残りとそれをサポートするデータと矛盾するかの検討は、残差逸 脱度、DIC、又はクロスバリデーションなどの標準的なモデル診断を利用して実行できる[1][5][25].

この方法の例は、記述的疫学[35][36][37][38]に、またスクリーニング応用でも見ることができる.早期型新生児B群溶血性連鎖球菌感染症(EOGBS: Early-onset group B Streptococcus infection)のモデルでは、ベースラインモデルに一連のパラメータが含まれていた:新生児GBS感染症の母親が保菌している確率、母親の保菌状態が与えられたもとでの新生児への伝播する確率、及び伝播が与えられたもとでのEOGBSの確率である.これらの確率のそれぞれに関する情報が入手可能である一方で、このモデルは、小児での臨床でのモニタリング計画[39]を通じて、イギリス諸島で報告されたEOGBSの症例数に関するデータに対して「キャリブレーション」された.この場合、本形式のキャリブレーションの効果は、個々の進行パラメータに非常に弱い制約を課すことであるが、それらの生成においては、かなり強い制約を課すことになる.このような方法は、長期間の追跡、疾患の登録、又は原因別死亡に関する独立したデータが利用可能な多数の臨床分野で、潜在的に適用される可能性があるが、明確な推奨がなされる前にさらなる研究が必要である.

参考文献

- [1] Dias, S., Welton, N.J., Sutton, A.J., Ades, A.E. NICE DSU Technical Support Document 2: A generalised linear modelling framework for pair-wise and network meta-analysis. 2011; last updated April 2012; available from http://www.nicedsu.org.uk
- [2] National Institute for health and Care Excellence. Guide to the methods of technology appraisal 2013.
- [3] Briggs, A., Claxton, K., Sculpher, M. Decision modelling for health economic evaluation. Oxford University Press, Oxford; 2008.
- [4] Barton, P., Bryan, S., Robinson, S. Modelling in the economic evaluation of health care: selecting the appropriate approach. Journal of Health Services & Research Policy 2004;9:110-118.
- [5] 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
- [6] Sculpher, M., Fenwick, E., Claxton, K. Assessing Quality in Decision Analytic Cost-Effectiveness Models. A Suggested Framework and Example of Application. Pharmacoeconomics 2000; 17(5):461-477.
- [7] Weinstein, M.C., O'Brien, B., Hornberger, J., et al. Principles of Good Practice for Decision Analytic Modeling

- in Health Care Evaluation: Report of the ISPOR task force on Good Research Practices-Modeling studies. Value Health 2003; 6:9-17.
- [8] Petrou, S., Gray, A. Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting. BMJ 2011; 342:d1766.
- [9] Golder, S., Glanville, J., Ginnelly, L. Populating decision-analytic models: The feasibility and efficiency of database searching for individual parameters. International Journal of Technology Assessment in Health Care 2005; 21:305-311.
- [10] Nuijten, M.J.C. The selection of data sources for use in modelling studies. Pharmacoeconomics 1998; 13:305-316.
- [11] Spiegelhalter, D., Thomas, A., Best, N., Lunn, D. WinBUGS User Manual Version 1.4 January 2003. Upgraded to Version 1.4.3. 2007; available from http://www.mrcbsu.cam.ac.uk/bugs
- [12] Hasselblad, V. Meta-analysis of multi-treatment studies. Med Decis Making, MED 1998; 18:37-43.
- [13] 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 2012; available from http://www.nicedsu.org.uk
- [14] Brown, J., Welton, N.J., Bankhead, C., Richards, S.H., Roberts, L., Tydeman, C. et al. A Bayesian approach to analysing the cost-effectiveness of two primary care interventions aimed at improving attendance for breast screening. Health Economics 2006; 15:435-445.
- [15] Welton, N.J., Ades, A.E., Caldwell, D.M., Peters, T.J. Research Prioritisation Based on Expected Value of Partial Perfect Information: a Case Study on Interventions to Increase Uptake of Breast Cancer Screening. Journal of the Royal Statistical Society (A) 2008; 171:807-841.
- [16] Govan, L., Ades, A.E., Weir, C.J., Welton, N.J., Langhorne, P. Controlling ecological bias in evidence synthesis of trials reporting on collapsed and overlapping covariate categories. Statistics In Medicine 2010; 29:1340-1356.
- [17]Rothman, K.J., Greenland, S., Lash, T.L. Modern Epidemiology. 3 ed. Lippincott, Williams & Wilkins, Philadelphia; 2008.
- [18] Burch, J., Paulden, M., Conti, S., Stock, C., Corbette, M., Welton, N.J. et al. Antiviral drugs for the treatment of influenza: a systematic review and economic evaluation. Health Technology Assessment 2010; 13(58):1-290.
- [19] Taylor, R.S., Elston, J. The use of surrogate outcomes in model-based cost-effectiveness analyses: a survey of UK health technology assessment reports. Health Technology Assessment 2009; 13(8).
- [20] Lu, G., Ades, A.E., Sutton, A.J., Cooper, N.J., Briggs, A.H., Caldwell, D.M. Metaanalysis of mixed treatment comparisons at multiple follow-up times. Statistics In Medicine 2007; 26(20):3681-3699.
- [21] Stettler, C., Wandel, S., Allemann, S., Kastrati, A., Morice, M.C., Schomig, A. et al. Outcomes associated with drug-eluting and bare-metal stents: a collaborative network meta-analysis. Lancet 2007; 370:937-948.
- [22]Nam, I.-S., Mengerson, K., Garthwaite, P. Multivariate meta-analysis. Statistics In Medicine 2003; 22:2309-2333.
- [23] Riley, R.D., Abrams, K.R., Lambert, P.C., Sutton, A.J., Thompson, J.R. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. Statistics In Medicine 2007; 26:78-97.
- [24] Riley, R.D., Thompson, J.R., Abrams, K.R. An alternative model for bivariate random effects meta-analysis

- when the within-study correlations are unknown. Biostatistics 2008; 9:172-186.
- [25] Ades, A.E. A chain of evidence with mixed comparisons: models for multi-parameter evidence synthesis and consistency of evidence. Statistics In Medicine 2003; 22:2995-3016.
- [26] Epstein, D., Sutton, A. Modeling correlated outcomes in health technology appraisals. Value in Health 2011; in press.
- [27] Welton, N.J., Cooper, N.J., Ades, A.E., Lu, G., Sutton, A.J. Mixed treatment comparison with multiple outcomes reported inconsistently across trials: evaluation of antivirals for treatment of influenza A and B. Statistics In Medicine 2008; 27:5620-5639.
- [28] Welton, N.J., Willis, S.R., Ades, A.E. Synthesis of Survival and Disease Progression Outcomes for Health Technology Assessment of Cancer Therapies. Res Synth Method 2010; 1:239-257.
- [29] National Institute for health and Care Excellence. Advanced Breast Cancer: diagnosis and treatment. NICE Clinical Guidelines, 2017.
- [30] Miller, D.K., Homan, S.M. Determining transition probabilities: confusion and suggestions. Med Decis Making 1994; 14:52-58.
- [31] Welton, N.J., Ades, A.E. Estimation of Markov Chain Transition Probabilities and Rates from Fully and Partially Observed Data: Uncertainty Propagation, Evidence Synthesis and Model Calibration. Med Decis Making, MED 2005; 25:633-645.
- [32] Price, M., Welton, N., Ades AE. Parameterisation of treatment effects for meta-analysis in multi-state Markov models. Statistics In Medicine 2011; 30:140-151.
- [33] Ades, A.E., Sutton, A.J. Multiparameter evidence synthesis in epidemiology and medical decision making: current appoaches. Statistics in Society, JRSS(A) 2006;169(1):5-35.
- [34] Vanni, T., Karnon, J., Madan, J., White, R.G., Edmunds, W.J., Foss, A.M. et al. Calibrating Models in Economic Evaluation: A Seven-Step Approach. Pharmacoeconomics 2011; 29:35-49.
- [35] Goubar, A., Ades, A.E., De Angelis, D., McGarrigle, C.A., Mercer, C.H., Tookey, P.A. et al. Estimates of human immunodeficiency virus prevalence and proportion diagnosed based on Bayesian multiparameter synthesis of surveillance data. Journal of the Royal Statistical Society A 2008; 171:541-580.
- [36] Presanis, A., De Angelis, D., Spiegelhalter, D., Seaman, S., Goubar, A., Ades, A. Conflicting evidence in a Bayesian synthesis of surveillance data to estimate HIV prevalence. Journal of the Royal Statistical Society (A) 2008; 171:915-937.
- [37] Sweeting, M.J., De Angelis, D., Hickman, D., Ades, A.E. Estimating HCV prevalence in England and Wales by synthesising evidence from multiple data sources: assessing data conflict and model fit. Biostatistics 2008; 9:715-734.
- [38] Presanis, A.M., De Angelis, D., Goubar, A., Gill, O.N., Ades, A.E. Bayesian evidence synthesis for a transmission dynamic model for HIV among men who have sex with men. Biostatistics 2011; 13:666-681.
- [39] Colbourn, T., Asseburg, C., Bojke, L., Phillips, Z., Claxton, K., Ades, A.E. et al. Prenatal screening and treatment strategies to prevent Group B Streptococcal and other bacterial infections in early infancy: cost-effectiveness and expected value of information analysis. Health Technology Assessment 2007; 11(29):1-226.

6 確率的な費用効果分析へのエビデンス統合の組み込み:ソフトウェアの選択 (TSD6)

NICE の TSD6[1]は確率的な費用効果分析へエビデンス統合を組み込む方法のガイダンスとして、各方法の留意点や用いられるソフトウェアの概要について記載された文書である. NMA 用のソフトウェアについては本 TF で追加調査を行い9章にまとめている.

意思決定において確率論的方法が提唱されている理由は主に二つある[2,3]. 一つ目は、期待される費用、便益、意思決定の不確実性における複数のパラメータの推定に伴う不確実性の影響が感度分析によって把握できるためである。これらの方法は確率的感度分析と呼ばれる。二つ目の理由は、パラメータ θ のベクトル/行列の不確実性の問題に直面したときに、意思決定者が、最も高い期待純便益(Expected Net Benefit)をもたらすオプション D*を複数のオプション D から選択できるためである。

$$D^* = \max_{D} E_{\theta}[NB(D, \theta)]$$

この "期待値" はパラメータ θ の同時分布における純便益の関数 $NB(D,\theta)$ の統合が必要である。統合方法の選択の適切性は、純便益関数の代数構造に依存しており、純便益がすべてのパラメータにおいて線形かつパラメータ間に相関がない場合を除いて、期待純便益はパラメータの期待値における純便益と同じではない。また、多くの費用効果分析(Cost Effectiveness Analysis: CEA)モデルにはマルコフモデルが含まれ、非線形であることが知られている。さらに、エビデンス統合では複数のパラメータを推定し、そのほとんどの場合でパラメータ間に相関がある。したがって、パラメータの複雑な不確実性構造が意思決定モデルに反映されるようなソフトウェアを用いるべきである。TSD6では全ての解析、つまり線形モデルであっても、モンテカルロシミュレーションベースのアプローチを推奨している。これらの分析において複数のソフトウェアを組み合わせる場合には、ツールを用いてソフトウェアを連携させることが可能である。

6.1 確率論的な費用効果分析へのエビデンス統合の組み込み

TSD6 では、エビデンス統合を確率論的な費用効果分析へ組み込む形として、ベイズ流で確率論的な費用効果分析まで 1 段階で行う方法(①)と 2 段階にわけて行う方法(②)、頻度論による推定をもとにモンテカルロ法を用いる場合(③)とブートストラップ法を用いる場合(④)について提示している.以下に各方法の特性と制限事項を以下に示す.

表 6-1 組み込み方法の特性の要約

方法の分類	推定方法	CEA 用ソフトへのアウトプット内容	制限事項
(1)	ベイズ流	なし: CEA は MCMC を実行したソフトで行	なし
(1)	MCMC	Ď	
2	ベイズ流	事後分布の標本(サンプリング)	なし
	MCMC		
2	ベイズ流	事後分布の平均,分散,相関係数	事後分布が多変量正規分布である
	MCMC		ことを仮定
2	ベイズ流	事後分布の平均,分散	パラメータ間に相関がないときの
	MCMC		み妥当*
3	頻度論	パラメータの推定値、分散共分散行列	治療効果が多変量正規分布に従う
			ことを仮定
3	頻度論	パラメータの推定値、分散	パラメータ間に相関がないときの
			み妥当*
4	頻度論	ブートストラップによるリサンプリング	スパースデータの場合は特別な方
			法が必要

*この方法を使う前に、ユーザーはデータ構造や解析方法がパラメータ間の相関を示唆していないことを確認すべきである.

(出典: TSD6 Table 1 一部改変・翻訳)

6.1.1 ベイズ流のシミュレーション:1 段階アプローチ

エビデンスの統合パラメータの推定に、事後分布からのサンプリングを用いるとき、単一のプログラミングのパッケージ(例えば WinBUGS や OpenBUGS)におけるプロセスとして CEA と統合することができる. これはいわゆる「包括的な意思決定分析 (Comprehensive Decision Analysis)」である.

このアプローチの利点は、事後分布からのサンプリングが意思決定分析に直接用いられることである. そのため、純便益やその他のアウトプットは、その分布についての仮定を必要とせずに、評価することができる.また、追加のパラメータなどを容易に組み込むこともできる.

MCMC のソフトウェアを適切に実行する知識がユーザーに必須というわけではないが、基礎的事項 やベイズ流データ分析への十分な理解は必要である. なお、MCMC 以外の方法として、標本再抽出のアルゴリズム(Re-sampling algorithm)[4]、ベイズ流のモデリング(Bayesian Melding)[5,6]、ベイズ流モンテカルロシミュレーション(Bayesian Monte Carlo)[7]などのベイズ更新スキーム提案されている. これら全てが、ベイズ推定と同時事後分布からのサンプリングという両方を特徴とする点で、MCMC の方法と同じ特性を持っており、特に後者二つはエビデンス統合のためにデザインされた方法である.

6.1.2 ベイズ流のシミュレーション:2 段階アプローチ

R, STATA, SASのような一般的なソフトウェアパッケージやEXCEL, TreeAGEといった決定樹 (decision tree) パッケージがあるCEA用ソフトウェアを所持している場合, MCMCなどで求めた事後 分布からのサンプリングを, CEA用ソフトウェアのインプットに使用するという選択がある. この方法ではパラメータの推定値の相関が保持されているため, 1段階アプローチと同じ特性を持つ.

6.1.3 モンテカルロサンプリングによる頻度論の推定

エビデンス統合が頻度論に基づくソフトウェアで行われた場合は、2段階アプローチが可能である. まず必要なパラメータとその分散共分散行列を推定し、2段階目として、他のCEAパラメータとともに モンテカルロサンプリングにそれを用いる. 従来の対比較のメタアナリシスのように、ひとつのパラ メータが重要である場合は、その分散だけを推定するので単純かつ容易である.

間接比較としては、A-B の比較試験、A-C の比較試験を別々に統合し、相対的な費用対効果を求めることは容易である(ただし共変量は考慮されていない).この方法も受け入れられはするものの、ベイズ流アプローチによる間接比較は共通の分散パラメータをもつ変量効果モデルに対応できる柔軟性を持っており、非常に少数の試験の比較の組がある場合には好ましいかもしれない.

6.1.4 ブートストラップによる頻度論の推定

ブートストラップはオリジナルデータからリサンプリングを繰り返し一連の新しいデータセットを作るテクニックであり、CEAでも用いられる.事後分布から得られたサンプリングと同じように扱うことができるものの、少数例やゼロセルが含まれる場合には単純にはいかない.ブートストラップのさらなる情報については他の文献を参照されたい[8,9,10].

6.2 複数のソフトウェアプラットフォームの使用

異なるソフトウェア・アプリケーション間でのデータ移送などのやりとりが可能なインターフェースを用いて、異なるソフトウェアで実施された CEA の構成要素を統合することができる. これは個々のパラメータだけでなく、異なるデータセットや異なる統合モデルでシナリオ分析を実施し、費用対効果の結果への影響をすばやく評価できる方法を提供することに繋がる.

このように統合されたアプローチを行う動機や利点は様々である。まず、異なるソフトウェアのスキルを持つ複数の専門分野のチームによる統合分析が可能なことが挙げられる。たとえば、統計家は一般的な統計ソフトウェアを使用し、意思決定者は EXCEL または特定の意思決定モデリングソフトウェアを使用することができる。また、解析の各コンポーネントに最適なソフトウェアを使用して、最適な結果を生成することもできる。例えば WinBUGS は NMA を実装するには最適なソフトウェアかもしれないが、作図機能は限られている。R のような高度な作図機能を有するソフトウェアで結果を提示することが望ましい場合がある。さらに、元のデータセットは、Excel などのスプレッドシートソフトウェアで用意されている可能性がある。

6.2.1 ソウトウェアの連携

ソフトウェアの連携方法について TSD6 で記載されているものをいくつか示す。また、ソフトウェアのコミュニケーション、透明性および将来のデータ更新のために、解析用に収集されたデータ(注釈および修正内容の詳細を含む)を 1 つのファイルとして保持しておくことが推奨されている。仮にWinBUGS で解析するのであれば、Excel などのスプレッドシートソフトウェアでデータ列をコピーし、WinBUGS のメニューの[Edit]から[Paste Special]を選択し、[Plain text]オプションを選択して貼り付けることができる。また、XL2BUGS は Excel データを WinBUGS ベクトル形式に変換する Excel アドインであり、BAUW はテキスト形式のデータを WinBUGS ベクトルまたは行列形式に変換する単体ソフトである。R にデータがある場合は、R の R2WinBUGS パッケージで WinBUGS にデータを渡すことができる。また、WinBUGS 実行のための SAS マクロ(SAS データを WinBUGS データに変換するマクロ、SAS による WinBUGS 実行マクロ、CODA データを SAS データに変換するマクロ)は製薬協データサイエンス部会 2014 年 TF5 の成果物として作成されている[11].

6.2.2 統合プラットフォームの使用

統合プラットフォームを使用することは、別のソフトウェアにデータや中間結果をコピーする際の手間や移送エラーのリスクを低減する. 他にも解析の修正や更新の容易化、感度分析の実施、透明性の促進といった利点もある.

統合プラットフォームの例として TSD6 で紹介されている Transparent Interactive Decision Interrogator (TIDI) [12]は、Excel 上で WinBUGS によるエビデンス統合および R による決定モデルの実施を可能にしたものであり、NICE の個別技術評価(STA:Single Technology Appraisal)向けに開発された。TIDI のコンセプトは、評価委員が会議中にリアルタイムで代替パラメータ値を使用して、CEA の再実行を可能にすることである。なお、TIDI に用いられている Excel から R を操作するツールである RExcel は本文書作成時点では有償ライセンスに変更されており、本 TF では動作確認はできていない。

参考文献

- [1] Dias, S., Sutton, A.J., Welton, N.J., Ades, A.E. NICE DSU Technical Support Document 6: Embedding evidence synthesis in probabilistic cost-effectiveness analysis: software choices. available from http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD6-Software.final_.08.05.12.pdf
- [2] Critchfield, G.C., Willard, K.E. Probabilistic analysis of decision trees using Monte Carlo simulation. Medical Decision Making 1986; 6(2):85-92.
- [3] Doubilet, P., Begg, C.B., Weinstein, M.C., Braun, P., McNeil, B.J. Probabilistic sensitivity analysis using Monte Carlo simulation: A practical approach. Medical Decision Making: an International Journal of the Society for Medical Decision Making1985; 5(2):157-177.
- [4] Rubin, D.B. Using the SIR algorithm to simulate posterior distributions. Bayesian Statistics 1988; 3:395-402.
- [5] Raftery, A.E., Givens, G.H., Zeh, J.E. Inference from a Deterministic Population Dynamics Model for

- Bowhead Whales. Journal of the American Statistical Association 1995; 90(430):402-416.
- [6] Poole, D., Raftery, A.E. Inference for Deterministic Simulation Models: The Bayesian Melding Approach. Journal of the American Statistical Association 2000; 95(452):1244-1255.
- [7] Brand, K.P., Small, M.J. Updating uncertainty in an integrated risk assessment: conceptual framework and methods. Risk Analysis 1995; 15(6):719-731.
- [8] Efron, B., Tibshirani, R.J. An introduction to the bootstrap. Chapman & Hall, New York; 1993.
- [9] Davison, A.C., Hinkley, D.V. Bootstrap methods and their application. Cambridge Univ Pr, 1997.
- [10] Lunneborg, C.E. Data analysis by resampling: Concepts and applications. Duxbury, 2000.
- [11]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. WinBUGS の使い方. 2014. http://www.jpma.or.jp/medicine/shinyaku/tiken/allotment/winbugs.html
- [12]Dr Sylwia Bujkiewicz University of Leicester https://www2.le.ac.uk/departments/health-sciences/research/biostats/staff-pages/sb309

7 意思決定における治療効果のエビデンス統合:レビューアー向けチェック リスト (TSD7)

7.1 チェックリストの想定利用者

ここで示すチェックリストは,有効性又は費用対効果に基づく意思決定で用いられる治療効果のエビデンス統合のレビュー用であり,間接比較やネットワークメタアナリシス(NMA)等による治療効果のエビデンス統合結果に対するレビューアーのチェック項目がリストされている.エビデンス統合を実施する申請者にとってもレビューアーの観点を理解しておくことは有用であると考えられるため,本チェックリストは,エビデンス統合を実施する申請者,エビデンス統合結果のレビューアーの両者に利用されることを想定している.

7.2 想定するチェックリストの利用場面

本チェックリストの利用対象は、事前規定された特性を伴う患者に対する治療の有効性又は費用対効果に関する比較検討を行うことを目的とする場合としている。体系的なエビデンスのレビュー及びメタアナリシスと関連する部分もあるが、保険償還の機関が活用でき、臨床家が関心をもつものとするため、意思決定に繋がるものに限定させている。また、エビデンス統合の質をランク付けすることが、意思決定のための妥当な基盤を有するかどうかの情報を与えるわけではないので、それを目的としているわけでもない。このチェックリストは、エビデンス統合の質に関わらず、存在するエビデンスが統合結果に公平に反映され結論づけられているかを評価するために利用されることが想定される。

7.3 チェックリストに含まれるもの

A.課題の定義

意思決定のための標的集団,対照薬群,試験の選択基準,治療の定義,試験結果,統合のための 測定尺度,標的集団に関連する試験での患者集団,試験の質,提示するデータ

- B. 解析方法と結果の提示 メタアナリシスの手法の記載、相対治療効果における異質性、試験結果についてのベースライ ンモデル、結果の提示
- C. ネットワーク統合に特化した課題 モデル特定, 多群試験, 連結ネットワークと分離ネットワーク, 不一致性
- D. 不確実性と相関の伝播を含む、確率的費用効果分析モデルでの統合

7.4 NMA における重要な仮定

本 TSD1~6 を通じ、対比較のメタアナリシスの拡張として NMA を定義している. この NMA で重要 となる仮定は、一連の試験で検討している治療のあらゆる組合せについて、その真の相対治療効果が固有のものか(固定効果モデル)または交換可能か(変量効果モデル)である.

7.5 対比較のメタアナリシスから NMA への拡張で追加となる課題

固定効果モデルまたは変量効果モデルにおける、相対治療効果が固有か、もしくは交換可能かについての要件は治療 X と Y のあらゆる組合せに対して存在する。それゆえ、対比較のメタアナリシスで必要とされる仮定に加え、NMA では「一致性」の仮定が必要となる。提案されているチェックリストの項目の多くは、対比較のメタアナリシスと NMA の両方に当てはまる。NMA では、ネットワークの連結、不一致性、そして対比較の実行が追加の課題となる。

7.6 他のチェックリストとの関係

これまでのガイドラインやチェックリストと比較すると、TSD1~6での提案は一部の領域で限定的な適用となるが、それ以外の領域ではそれほど限定的にならない。限定的な適用になる領域は、効果修飾因子の扱いが必要な領域である。臨床的に明確に定義された同質な標的集団に注目することが推奨されるのは、異質性を減らし不一致性のリスクを減少させるためではなく、意思決定の課題に関係している可能性が高いためである。これ以外では、提案されているアプローチは限定的なものではない。ここで提案されているチェックリストは、ISPORのタスクフォースのチェックリスト[1][2]に類似した項目を含んでいるが、学術的な論文投稿のためではなく、実際の意思決定プロセスに適応した構成になっている。

7.7 チェックリストを提供する目的

チェックリストを提供する目的は、エビデンス統合のレビューアーがどのような質問をすべきかについてのガイダンスを与えることである。レビューアーは記載の明確化、感度解析や探索アルゴリズムの詳細、コンピューターコードなどを要求する可能性があると想定している。チェックリスト自体の目的は、エビデンス統合の申請でどのような仮定を置いているか、どのような経験的エビデンスや事実に基づき理由づけられているか、改訂の申請では課題が解決されることを確認するためにどのようなステップが取られているか、等のレビューアーの正確な理解を助けることである。

そのため、このチェックリストは、エビデンス統合の実施及び仮定についての事実の記録として、コメント欄も設けられている。仮定またはエビデンスの解釈が疑わしい場合は、追加解析もしくは感度解析の必要性が指摘される可能性がある。場合によっては、エビデンスの欠落のためより強い仮定が必要

になる可能性がある. さらに、これらの仮定を確認する経験的なアプローチは利用可能なデータにより制限される可能性がある. 全ての仮定とそれらの結果への関わりの一貫した透明性のある議論が、その報告書に盛り込まれるべきである. このチェックリストは、レビューアーが、仮定が妥当で適切に説明されるか否かを確認すること、そして疑問点に適切に対応することに考慮して構成されている.

7.8 チェックリスト

		十分か?	コメント		
A. 課是	便の定義				
A1. 意	思決定のための標的集団				
A1.1	意思決定のための標的患者集団は明確に定義されているか?				
A2. 比	較対照				
A2.1	意思決定の比較集団:意思決定における適切な治療全てが特定されているか?				
A2.2	統合比較集団:意思決定の比較集団以外に,統合比較集団における追加の治療はあるか. その場合,適切に説明されているか?				
A3. 試	・ 験の選択基準/除外基準				
A3.1	検索方法は技術的に妥当で適切に報告されているか?				
A3.2	統合する比較対照集団として、少なくとも2治療を含んでいる試験全 てが含まれているか?				
A3.3	関連する結果を報告している試験は全て含まれているか?				
A3.4	追加の試験は含まれているか?そうであれば、これは適切に説明されているか?				
A4 治线					
A4.1	全ての治療オプションが特定の用量、併用治療について制限されているか. もしくは異なる用量と併用治療は「ひとかたまり」に扱われているか? 後者の場合、適切に説明されているか?				
A4.2	モデリングの追加の仮定は何かあるか?				
A5 統台	A5 統合のために選択された試験結果と測定尺度				
A5.1	他の結果尺度が利用可能なとき,統合で利用される結果尺度の選択は 説明されているか?				
A5.2	尺度の選択の背後にある仮定は説明されているか?				
A6 患者集団:標的集団以外の患者の試験					
A6.1	標的集団以外の患者の試験は含まれるか?そうであれば,これは適切に説明されているか?				
A6.2	どのような仮定が相対治療効果の大きさに影響を及ぼすか,もしくは 影響を及ぼさないか?それらは適切に説明されているか?				

		,				
	患者集団の違いのための調整がされているか?					
A6.3	調整されている場合、調整の下で提示されるエビデンスの妥当性、感					
	度解析の必要性についての考察が必要である.					
A7 患者	f集団:標的集団の中での異質性					
A7.1	治療効果の潜在的な修飾因子は考慮されているか?					
A7.2	標的集団内であっても,試験間で患者集団の明確もしくは潜在的な違					
A7.2	いはあるか?そうであれば、これは適切に考慮されているか?					
A8 バ	イアスのリスク					
A8.1	これらの試験,もしくは試験の全体的効果が脆弱になりうるバイアス					
A6.1	の議論はあるか?					
A8.2	バイアスのリスクが特定されている場合,解析で調整され,適切に説					
A0.2	明されているか?					
A9. デ	ータの提示					
A9.1	どのようなデータが基準症例(base-case)の解析に含まれているかを示					
A9.1	す、明確な表やグラフはあるか?					
A9.2	どのようなデータがどのような理由で除外されているかを示す,明確					
A9.2	な表やグラフはあるか?					
B. 解析	f手法と結果の提示					
B1 ≯ 3	タアナリシスの手法					
B1.1	統計モデルが明示されているか?					
B1.2	ソフトウェアの実行は文書化されているか?					
B2. 相	対治療効果の異質性					
DO 1	相対治療効果における異質性の程度について, 定量的な推定が与えら					
B2.1	れているか?					
B2.2	変量効果モデル・固定効果モデルからのモデル選択は説明されている					
D 2.2	か?感度解析も実施されるべきではないか?					
B2.3	異質性に対する妥当な回答はあるか?					
B2.4	相対治療効果において、結論の頑厳性を脅かす説明不可能な変動はあ					
D2.4	るか?					
B2.5	ベースラインの群間差の統計的な異質性は議論されているか?					
B3 試懸	B3 試験結果に対するベースラインモデル					
	ベースライン効果と相対治療効果は同じモデルで推定されている					
B3.1	か?					
	その場合,それは説明されているか?					
D2 2	ベースラインモデルを満たすように利用された試験の選択は説明さ					
B3.2	れているか?					

B4 試験	・ データの解析結果の提示			
= · # (4)	(プラセボまたは標準対照薬への) 相対治療効果は、変量効果モデル			
B4.1	が利用されるときは試験間の異質性の観測値と共に表で与えられて			
	いるか?			
	費用効果分析で利用されている各治療の絶対効果が報告されている			
B4.2	カゝ?			
B5 自然	病歴 (natural history)モデルの他の部分における統合			
D.F. 1	自然病歴モデルにおいて,他のパラメータを特徴づけるためのデータ			
B5.1	ソースの選択は,適切に記載され説明されているか?			
B5.2	自然病歴モデルにおいて、治療間の長期投与の違いを無作為化試験の			
D 3.2	結果の違いから説明可能か?			
C. ネッ	トワーク統合に特化した課題			
C1 モラ	デル特定における情報の妥当性とソフトウェア実行			
C2. 多	詳試験			
C2.1	多群試験があるとき,相対治療効果の間の相関は考慮されているか?			
C3 連綿	C3 連結されたネットワークと分離されたネットワーク			
C2 1	無作為化試験に基づくエビデンスのネットワークは連結されている			
C3.1	C3.1 // // ?			
C4 不一	- 致性			
C4.1	ネットワークの中でいくつの不一致を含められるか?			
	治療 A と治療 B を比較する試験の患者と治療 A と治療 C を比較する			
C4.2	試験の患者の間の体系的な臨床的特性の違いにより,不一致が存在す			
	るかもしれないことについての事前の理由はあるか.			
C4.3	不一致性の適切な点検が行われているか?			
D 確 ^図	D 確率的費用効果分析における統合の構成			
D1. 不	D1. 不確かさの拡大			
D1.1	パラメータ推定における不確かさは費用効果分析モデルを通じて拡			
D1.1	大されているか?			
D2 相関	D2 相関関係			
D2.1	パラメータ間の相関はあるか?相関がある場合は,費用対果分析モデ			
D2.1	ルを通して伝播されているか?			

参考文献:

- [1] Jansen, J.P., Fleurence, R., Devine, B., Itzler, R., Barrett, A., Hawkins, N. et al. Interpreting Indirect Treatment Comparisons and Network Meta-Analysis for Health-Care Decision Making: Report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: Part 1. *Value in Health* 2011; 14:417-428.
- [2] Jansen, J.P., Fleurence, R., Devine, B., Itzler, R., Barrett, A., Hawkins, N. et al. Conducting Indirect Treatment Comparisons and Network Meta-Analysis Studies: Report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: Part 2. *Value in Health* 2011; 14:429-437.

8 NICE DSU TSD の WinBUGS プログラムについて

8.1 プログラムの紹介

各 TSD の付録に統合対象の試験群の説明とコードが記載されているものの、付録のコードには一部 誤記がある. そのため、以下の webs サイトからダウンロードできる WinBUGS system (.odc) files を閲覧 することを推奨する. このファイルにはコードだけでなく結果も格納されているため、TSD に転記されているパラメータ以外についても確認することができる. 本 TF ではこれらのファイルが WinBUGS で再実行できることを確認している. なお、OpenBUGS では一部のファイルは開くことができないため注意されたい.

Evidence Synthesis TSD series

 $\underline{http://nicedsu.org.uk/technical-support-documents/evidence-synthesis-tsd-series/}$

(Last access date: 2020/02/20)

NICE DSU TSD2-5 で公開しているプログラムリスト:

Evidence synthesis TSD series			
	A general linear modelling framework for pair-wise and network meta-analysis of randomised controlled		
TSD2	trials (last updated Sept 2016)		
WinBUGS system(.odc) files (last updated Sept 2016)			
Heterogeneity: subgroups, meta-regression, bias and bias-adjustment			
TSD3	WinBUGS system (.odc) files		
TSD4 Inconsistency in networks of evidence based on randomised controlled trials (last updated WinBUGS system (.odc) files (last updated March 2013)			
		TODE	Evidence synthesis in the baseline natural history model
TSD5	WinBUGS system (.odc) files		

プログラムの説明:

TOD	統合対象の	プログラムで扱ったモデルの説明:特に記載のない場合,
TSD	試験群の略称	群単位の治療効果の要約指標に対する NMA
TSD2	BLOCKER	Binomial likelihood, logit link, 変量効果モデル, 治療法が二つの場合のみに対応
		/固定効果モデル
		Binomial likelihood, logit link,変量効果モデル/ 固定効果モデル
	DIETARY FAT	Poisson likelihood, log link, 変量効果モデル/ 固定効果モデル
	DIABETES	Binomial likelihood, cloglog link,変量効果モデル/ 固定効果モデル
	SCHIZOPHRENIA	Multinomial likelihood (with competing risks), log link, 変量効果モデル/ 固定効果
		モデル
	PARKINSON'S	Normal likelihood, identity link,変量効果モデル/ 固定効果モデル

	PSORIASIS	Conditional Binomial likelihood, probit link,変量効果モデル/ 固定効果モデル
	PARKINSON'S	Normal likelihood, identity link, 変量効果モデル,群間差の要約指標に対するモデル/固定効果モデル
	PARKINSON'S	Normal likelihood, identity link, 変量効果モデル,群単位と群間差の要約指標が混在した shared parameter model /固定効果モデル
TSD3	MAGNESIUM	Binomial likelihood, logit link, 変量効果モデル,
1523	WITGI (ESTEN)	ペアワイズメタアナリシスにおける予測的クロスバリデーションの実施
	Adverse events in	Binomial likelihood, logit link, 変量効果モデル, 予測的クロスバリデーションの
	Chemotherapy	実施
	STATINS	Binomial likelihood, logit link, 変量効果モデル, サブグループ効果があるメタ回
		帰/固定効果モデル
	BCG VACCINE	Binomial likelihood, logit link, 変量効果モデル, 連続量の共変量のあるメタ回帰/
		固定効果モデル
	CERTOLIZUMAB	Binomial likelihood, logit link, 変量効果モデル, 連続量の共変量のあるメタ回帰/
		固定効果モデル/変量効果の分散を informative prior distribution とした場合
		注:TSD3 例 4 を基にしているプログラムのため,ファイル名は TSD3-
		4aRE_continuousCZP, TSD3-4bRE_continuousCZP となっている
	CERTOLIZUMAB	Binomial likelihood, logit link, 変量効果モデル, ベースラインリスクを調整する
		メタ回帰/固定効果モデル/変量効果の分散を informative prior distribution とした
		場合
TSD4	SMOKING	Binomial likelihood, inconsistency model,変量効果モデル
	CESSATION	
	THROMBOLYTIC	Binomial likelihood, inconsistency mode,固定効果モデル
	TREATMENTS	
TSD5	SMOKING	Binomial likelihood,ベースライン変量効果モデルと予測分布
	CESSATION	
	SMOKING	Binomial likelihood, ベースラインと治療効果の変量効果モデルと予測分布
	CESSATION	

8.2 WinBUGS の基本的な使い方の補足

WinBUGS の基本的な使い方については製薬協データサイエンス部会 2014 年 TF5 が作成した「WinBUGS の使い方」[1]が参考になる. ここではこの資料の補足として,使用するデータセットの設定とイタレーション(反復)の破棄がある場合について説明する.

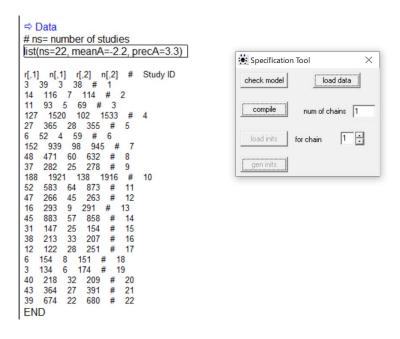
8.2.1 使用するデータセットの設定について

WinBUGS は、データのフォーマットとしてリストフォーマット及び矩形フォーマットを取り込むことができる。リストフォーマットの構文には、"list"から開始し、データを丸括弧で囲み、括弧内に各変

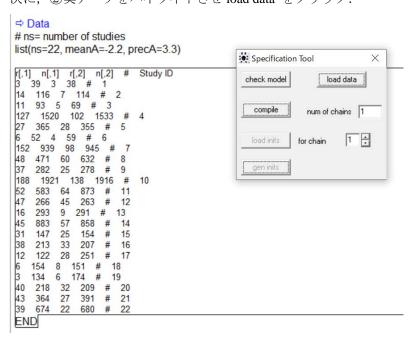
数:試験数 (ns), 平均値 (meanA), 推定値 (precA) をカンマによって区切って入力する. 各治療群における有効症例数 (分子) を r[,1], 総症例数 (分母) を n[,1]に入力, #で区切り Study ID などを入力する.

下記の画面で入力されたデータ⇒Data を読み込むには2回選択する必要がある.

① 最初は「list」の文字だけハイライトさせ load data をクリック.



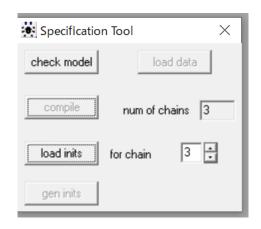
次に、②実データをハイライトさせ load data をクリック.



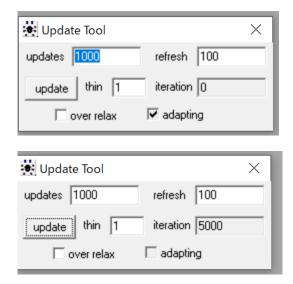
問題無く読み込めていれば WinBUGS の画面下に「data loaded」と表示される.



8.2.2 初期値の影響があるイタレーション(反復)の破棄がある場合について



初期値の影響があるイタレーション(反復)の破棄がある場合,以下のように"gen inits"がクリックできないため,メニューバーから model \rightarrow Update で[Update Tool] を開く必要がある.その後,update をクリックし[iteration]が 1000 になるまで待ち,1000 回では,まだ初期値の影響があり定常状態に達していない場合,[adapting]の状態になる.もう一度 Update をクリックして[adapting]のチェックが外れるまで繰り返す必要がある.(参照 Section 4.2.4 p69 The Bugs book)[2]



参考文献:

- [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(2014). WinBUGS の使い方. http://www.jpma.or.jp/medicine/shinyaku/tiken/allotment/winbugs.html
- [2] The BUGS Book: A Practical Introduction to Bayesian Analysis1st Edition David Lunn, Chris Jackson, Nicky Best, Andrew Thomas, David Spiegelhalter

9 NMA を実施するソフトウェアの紹介

本章では本 TF で調査した Network meta-analysis (NMA)を実施するためのソフトウェアを紹介する. ここで紹介するいくつかのソフトウェアについては、その機能を比較している文献[1][2]があるものの、本章での紹介文と同じく調査時の情報であるため、実際にソフトウェアを用いる際には、各ソフトのサポートページ等で最新の情報を確認していただきたい、ソフトウェアは将来的に使えなくなる可能性もあり、特に複数のソフトウェアを連携するタイプのものには注意が必要である.

9.1 NMA を実施するソフトウェアの分類

NMA を実施するソフトウェアは以下のように分類でき、主なソフトウェアについて次節以降で紹介する

● MCMC 専用ソフトウェア

ベイズ流の NMA を実施するために必要なマルコフ連鎖モンテカルロ (Markov Chain Monte Carlo; MCMC) 法の専用ソフトウェアであり、自分でモデルを記述する. SAS などの汎用の統計解析ソフトウェアに比べてモデルを記述しやすいが、データハンドリングや図示機能が劣るため、他と組み合わせて使う場合が多い.

● NMA 専用のソフトウェア

専用設計されているため扱いやすく、GUI対応やネットワークダイアグラムの図示機能もあることが多い. しかしながら組み込まれているモデルのみ利用可能となることが多い. 実際の解析は WinBUGS などの MCMC 専用ソフトを利用している場合もある.

● 汎用の統計解析ソフトウェア

汎用機能または MCMC 専用ソフトウェアを用いて自分でモデルを作成する方法と、アドインとして NMA 用の機能を追加できる場合がある. 後者の場合は NMA 専用ソフトウェアと同様なメリットとデメリットがある.

9.2 MCMC 専用ソフトウェア

WinBUGS

MCMC 法による解析を行う Windows 用のソフトウェアであるが、v1.4.3 で開発終了している. Windows10 でも使用できるものの正式サポートはされていない. 代替ソフトウェアとして、同じ BUGS 言語による OpenBUGS や JAGS が候補となり、基本的には WinBUGS で書かれたコードで 実行できる. なお、Windows10 でインストールするにはインストーラー版(exe ファイル)ではなく zip ファイル版を用いる必要がある. 本文書で紹介している NICE DSU TSD を始めとして、豊富な

事例・実装例があり、NMA の方法論の研究論文やシステマティックレビューに BUGS 言語のコードがついている場合もある. TSD の付録にあるプログラムを動かす際には製薬協データサイエンス部会 2014 年 TF5 の成果物[3]および本文書の 8 章に示した WinBUGS の使い方の補足を参照されたい.

WinBUGS https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/

OpenBUGS http://www.openbugs.net/w/FrontPage

JAGS http://mcmc-jags.sourceforge.net/

• Stan (https://mc-stan.org/)

NUTS(No-U-turn sampler)アルゴリズムに基づいたハミルトニアン・モンテカルロ法(HMC)を採用している. また, WinBUGS に比べてデバッグしやすいといわれている.

9.3 NMA 専用のソフト

• NetMetaXL (http://www.NetMetaXL.com)

WinBUGS で MCMC を行うものの、Excel ベース(VBA)で作成されているため、プログラミング に不慣れでも実行できる。NICE TSD で検討されたベイズ流の NMA を GUI で実行でき、WinBUGS のスクリプトや log も標準で出力される. また, カナダの医療技術評価機関 CADTH (Canadian Agency for Drugs and Technologies in Health)が協力している.

- ▶ 生成される図表:ネットワークダイアグラム、フォレストプロット、相対治療効果のリーグ テーブル、各治療の順位確率の積み上げ棒グラフ、一致性を仮定したモデルと一致性を仮定し ないモデルの相対治療効果の散布図
- ▶ v1-6-1 で対応している統合対象データの制限は以下のとおり
 - ♦ 解析データセット:治療法の数は15,試験数は50まで
 - ◆ 評価項目:2値データ
- ▶ 使用方法の説明動画の URL(CADTH 公式): https://www.youtube.com/user/CADTHACMTS

GeMTC

本文書で紹介した NICE TSD で検討されたベイズ流の NMA が実施可能であり、統計解析ソフト R に同じ作者によるパッケージ gemtc がある. 開発は停止しており、現在は web 上で実施する形式の ADDIS 2 へ移行している.

GeMTC(非 R パッケージ) https://drugis.org/software/addis1/gemtc

ADDIS2 https://drugis.org/software/addis2/index

9.4 汎用の統計解析ソフトウェア

R

R の NMA 用パッケージ(ライブラリー)は以下のとおり複数公開されている。それぞれの機能や統合対象の試験データセットの構造はパッケージによって異なるため、各ヘルプを参照されたい。また、NMA 専用パッケージを使わずに MCMC によるベイズ流の NMA を行うためには、MCMC 汎用パッケージ MCMCpack を用いる以外にも、MCMC 専用ソフトを利用する方法がある。連携用のパッケージとしては、WinBUGS (JAGS, OpenBUGS)に対応する R2WinBUGS (rjags/R2jags/runjags, BRugs)、Stan に対応する Rstan がある。

RのNMA用パッケージ:

- ☐ gemtc: Network meta-analyses (mixed treatment comparisons) in the Bayesian framework using
- 本文書で取り扱った NICE TSD のベイズ流の NMA が実行でき、いくつかのサンプルデータも 組み込まれている. ver0.7 からは外部の MCMC ソフトウェアとして JAGS のみに対応(RJGAS 利用)となり、WinBUGS と OpenBUGS には非対応となったが、residual deviance statistics の追加 とともに DIC 算出方法が NICE TSD 流へ変更されている(TSD2 参照)
- □ netmeta: Network Meta-Analysis using Frequentist Methods
- 頻度論に基づく NMA が利用できる. 2020 年 1 月 15 日時点の最新バージョンは 1.2-0 (January 9, 2020)と本文書作成時も開発が続いている.
- ☐ Pcnetmeta: Performs arm-based network meta-analysis for datasets with binary, continuous, and count outcomes using the Bayesian methods
- □ nmaINLA: Network Meta-Analysis using Integrated Nested Laplace Approximations
- BUGSnet: an R package to facilitate the conduct and reporting of Bayesian network Meta-analyses

SAS

SAS には NMA 専用の機能はないものの、ベイズ流または頻度論によるいくつかの実装例が SAS ユーザー総会等で報告されおり、本 TF も SAS プログラムを公開している[4][5][6][7]. なお、SAS から WinBUGS を利用する場合は、製薬協データサイエンス部会 2014 年 TF5 の成果物[3]で提供されているマクロが有用である.

STATA

STATA には NMA 用機能があり、頻度論に基づく一連の方法が mvmeta で実施できる. また、WinBUGS や stan を用いることも可能である.

Python

Python は SAS 等と比べてより汎用的なプログラミング言語ではあるが, MCMC 機能として, PyStan (Stan 利用), PyJAGS (JAGS 利用), PyMC(Python の MCMC パッケージ)があるため, NMA も実施可能である.

参考文献:

- [1] Neupane, B., Richer, D., Bonner, A. J., Kibret, T., & Beyene, J. (2014). Network meta-analysis using R: a review of currently available automated packages. *PloS one*, 9(12), e115065.
- [2] Xu C., Niu Y., Wu J., Gu H., Zhang C. (2018). Software and package applicating for network meta-analysis: A usage-based comparative study. J Evid Based Med. Aug;11(3):176-183.
- [3] 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(2014). WinBUGS の使い方. http://www.jpma.or.jp/medicine/shinyaku/tiken/allotment/winbugs.html
- [4] 福井 伸行, 乙黒 俊也, 磯崎 充宏. ネットワークメタアナリシスによる無作為化比較試験の統合. SAS ユーザー総会 2014
- [5] 舟尾暢男, 黒田晋吾. SAS での Network Meta-Analysis の実施例 〜頻度論に基づくアプローチ〜. SAS ユーザー総会 2016
- [6] 奥山ことば,町田 光陽, 石渡 量太, 渥美 淳, 奥村 広之. 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. SAS ユーザー総会 2019
- [7] Madden, L. V., Piepho, H. P., & Paul, P. A. (2016). Statistical models and methods for network meta-analysis. Phytopathology, 106(8), 792-806.

おわりに

本報告書では、NICE DSU が公開した TSD のエビデンス統合の基礎的な内容を含めた TSD 1-7 をもとに、一般化線形モデルのフレームワークによる NMA を紹介し、治療の相対治療効果を正確に評価するために必要な、外れ値の検討、メタ回帰及びバイアスの調整、不一致性の考え方及び評価方法、ベースラインモデル、費用効果分析にエビデンス統合を組み込む際のソフトウェア選択の考え方、レビューアー用のチェックリスト、さらには NMA を実施するためのソフトウェアおよび TSD 付録の WinBUGSファイルの使い方の補足説明をまとめた。なお、その他のエビデンス統合に関わるアドバンス的な内容を含んだものとして TSD18(個別データの利用)及び TSD20(複数アウトカムのメタ解析)があり、これらは 2020 年度 TF6 にて検討を行う予定である。

冒頭でも触れた通り、NMA は複数の治療法を比較できるという非常に魅力的な特徴をもつ反面、その結果の妥当性を担保するために満たすべき仮定も多いため、NMA を活用する際には解析結果およびそこに至る過程を十分に吟味し、慎重に解釈すべきである。特に、臨床疫学的知見およびシステマティックレビューによる十分な検討が必要となることは再度強調されるべきであろう。本報告書が費用対効果評価における NMA の理解にわずかでも役立てば幸いである。

2019年度データサイエンス部会KT5

作成担当者

渥美 淳 東レ株式会社 担当:4章,6章,9章

石渡 量太 ユーシービージャパン株式会社 担当:1章,8章

奥山 ことば MSD株式会社* 担当:3章,5章

町田 光陽 塩野義製薬株式会社 担当:2章,7章

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

直井 一郎 大日本住友製薬株式会社

野島 俊秋 興和株式会社

松下 泰之 第一三共株式会社*

吉田 瑞樹 ファイザーR&D合同会社

*TFリーダー

担当副部会長

酒井 弘憲 エーザイ株式会社

データサイエンス部会レビューワー

兼山 達也 大日本住友製薬株式会社

酒井 弘憲 エーザイ株式会社

藤井 陽介 ファイザーR&D合同会社

山本 英晴 中外製薬株式会社