

Evidence Synthesis
and Meta-Analysis
for Drug Safety

医薬品安全性における
エビデンスの統合
とメタアナリシス

CIOMS Working Group X 報告

Geneva 2016

仮訳：海外情報分科会

目次

はじめに	vi
第1章 概要とイントロダクション	1
1.1 主題	1
1.2 概要と範囲	2
1.3 読者層	5
1.4 将来	6
第2章 背景	7
2.1 本書を取りまく状況	7
2.1.1 現在の研究統合の背景	8
2.2 医薬品開発と規制プロセスを取りまく状況	10
2.3 なぜエビデンスの統合が必要か	15
2.4 研究の統合における重要な留意点	15
2.5 結論	17
第3章 計画	18
3.1 メタアナリシスのための準備	18
3.2 対象母集団の定義	23
3.3 アウトカムの定義	25
3.3.1 前向きメタアナリシスのアウトカム定義	25
3.3.2 後ろ向きメタアナリシスのアウトカム定義	27
3.4 異なる期間の試験	29
3.5 研究の選択とデータの抽出	31
3.5.1 研究の選択	32
3.5.2 データの抽出	36
3.6 研究規模と小規模研究効果	37
3.7 多群試験	40

3.8	メタアナリシスのバイアスリスク、個々の研究と研究全体.....	42
3.8.1	早期試験中止によるランダム化の脆弱性.....	43
3.8.2	割り当ての非盲検化.....	44
3.8.3	有害事象の確認.....	45
3.8.4	個別試験のデータ収集.....	46
3.8.5	併合のための試験の適合性.....	48
3.8.6	分析単位の問題.....	49
3.9	個別参加者データ、要約レベルデータまたは両者へのアクセス.....	50
3.9.1	個別参加者レベルデータの利点.....	51
3.10	観察研究のメタアナリシスにおける特有の問題.....	54
3.10.1	背景.....	54
3.10.2	原則.....	55
3.10.3	観察研究データの妥当性に対する問題.....	57
3.10.4	観察研究を用いたメタアナリシスの結論.....	59
3.11	ネットワークメタアナリシス.....	60
第4章	解析とレポート作成.....	63
4.1	治療効果の測定.....	63
4.2	希少事象のメタアナリシスのための統計手法.....	67
4.3	統計モデルの選択.....	71
4.3.1	固定効果モデルとランダム効果モデル.....	71
4.3.2	モデル特定/分析的手法.....	74
4.3.3	パラメータ推定.....	75
4.3.4	治療効果の統計学的異質性.....	76
4.4	ベイジアンメタアナリシス.....	77
4.4.1	ベイズ法の長所と短所.....	78

4.4.2	追加情報.....	79
4.4.3	研究の統合における重要な留意点	80
4.5	検出力 (Power)	81
4.6	欠測値.....	83
4.6.1	欠測値の概要.....	83
4.6.2	メタアナリシスにおける欠測値の取り扱いに一般的に用いられる方法.....	85
4.7	メタアナリシスにおける多重性と累積メタアナリシス	86
4.8	異質性.....	90
4.8.1	サブグループ解析.....	93
4.8.2	メタ回帰分析.....	94
4.8.3	サブグループ解析とメタ回帰分析に関する総論	95
4.9	感度分析.....	97
4.10	メタアナリシスの報告.....	99
4.10.1	フォレストプロットの使用.....	99
4.10.2	報告のためのチェックリスト	101
第5章	結果の解釈.....	109
5.1	序論	109
5.2	メタアナリシスの潜在的な臨床上の重要性	110
5.3	メタアナリシスの技術的妥当性	111
5.3.1	メタアナリシスのための計画の評価	112
5.3.2	正しくない解答を得る可能性 (バイアスの可能性、対象集団への調査結果の適応性を含む)	115
5.3.3	多重性.....	116
5.3.4	治療効果の異質性	118
5.4	結果の頑健性.....	119
5.4.1	研究のデザインに関する懸念事項.....	120

5.4.2	エンドポイントの臨床的妥当性に関する懸念事項	123
5.4.3	解析自体に関する懸念事項	124
5.5	エビデンスの統合と解釈	126
5.5.1	安全性懸念事項や患者への影響における臨床、医学上の妥当性	127
5.5.2	安全性懸念事項の研究におけるメタアナリシスの重要性の重み付け	130
5.5.3	安全性の懸念事項を取り巻く状況	131
5.6	医薬品の使用に与える影響	132
5.7	結果報告における注意点	135
5.8	結果のコミュニケーションにおける留意点	136
5.8.1	解析の異なる段階におけるコミュニケーションの留意点	137
5.8.2	メタアナリシスの情報を組み込んだベネフィット・リスク評価の留意点	138
5.8.3	メタアナリシスとベネフィット・リスク評価から生じるガイダンスの留意点	139
5.8.4	著者、ジャーナル編集者と規制当局の関係の留意点	140
5.9	結び	140
第6章	リソース	141
6.1	コアチーム – メタアナリシス実施の組織的側面	141
6.2	一般的リソース	142
6.3	検索戦略およびデータベース	144
付録1	事例研究 – 成人における抗うつ薬と自殺	148
1.	背景	148
2.	メタアナリシスの概要	149
3.	具体的対応	151
付録2	事例研究 – 抗腫瘍領域における赤血球造血刺激因子製剤（ESA）と血栓塞栓症	153
1.	背景	153

2.	メタアナリシスの概要	154
3.	具体的対応	155
付録 3.	事例研究－COPD 患者におけるチオトロピウム及び心血管イベント	157
1.	背景	157
2.	メタアナリシスの概要	157
3.	具体的対応	158
付録 4.	事例研究－糖尿病患者におけるロシグリタゾンと心血管イベント	160
1.	背景	160
2.	メタアナリシスの概要	161
3.	具体的対応	162
補遺 I.	用語集.....	164
補遺 II.	用語集の事例研究－ホルモン補充療法及び乳がんにおける効果測定.....	172
補遺 III.	CIOMS X ワーキンググループメンバー及び会議	174
BIBLIOGRAPHICAL REFERENCES.....		177

はじめに

医薬品は、承認時に、ベネフィット（有効性）とリスク（副作用）のバランスが良好であることが期待されている。しかし、製品のライフサイクルにおいて、開発中に特定されたリスクの特徴を明確化する必要があるとともに、開発中に認識できなかった潜在的なリスクが製造販売後に出現することがある。既存の医薬品においても、開発中や発売後数年では認識できなかった安全性の問題を徹底的に調査することが必要になる場合もある。

医薬品の特定されたリスクや潜在的なリスクを評価し、安全性の問題を厳密に追及する目的で、過去数十年にわたる承認後の医薬品安全性監視（**pharmacovigilance**）活動を進める中で各種の規制ツールが開発されてきた。医薬品の全ライフサイクルで潜在的なリスクの評価に使用される科学的ツールとして、メタアナリシスとシステマティック・レビューの利用が増加している。特定の治療法に伴う有害事象の有無を明らかにするために、蓄積されたデータを用いて安全性に関するメタアナリシスが実施されている。通常の目的は、将来の患者保護のため、手元にある研究結果の平均効果を推定することである。

CIOMS ワーキンググループ（WG）は、規制当局、製薬企業、関連学界等の専門家から構成されており、医薬品の安全性に関する幅広いトピックを扱ってきた。各WGは当初から、開発時と市販後における医薬品の潜在的なリスクの検出とその管理プロセスに焦点を当ててきた。重大な予期せぬ副作用の個別事例について「CIOMS フォーム」による標準化された有害事象の国際報告のための最初のガイドラインを 1990 年に発表し、続いて定期的安全性最新報告書（PSUR）や中核データシート（Core Data Sheet）についてのガイドラインを発表した。

当初のWGは承認後のプロセスに集中したが、時代が進んで新しいWGが構成されると、その範囲は、医薬品のライフサイクル全体の安全性に拡大した。CIOMS WG Vでは「ファーマコビジランスー市販後監視への新しい取り組み」、VIでは「臨床試験からの安全性情報の取り扱い」、VIIでは、「開発段階における定期的安全性最新報告」と題し、臨床試験中の定期安全性報告書の書式と内容の調和などの分野をカバーしている。これに続いて、CIOMS WG VIIIでは「ファーマコビジランスにおけるシグナル検出の実践」と題し、安全性シグナルの検出、優先順位付け、評価を含むシグナルのライフサイクルを管理するための報告書を刊行した。

直近のCIOMS WG IXの報告書では、医薬品のリスク最小化のための実践的アプローチを提示した。

CIOMS WG X の活動では、WG メンバーの一部が在籍する組織や企業のさまざまな施設で、メタアナリシスとシステマティック・レビューの話題について多くの会合を開催した（補遺 III 参照）。メンバーは合意達成に向け可能な限りの努力をし、本書に記載されている見解に対し、専門家として責任を負う。しかしながら、これらの見解は、必ずしもメンバーの所属する機関または企業の見解あるいは政策や意思決定を意味するものではない。

本書には、特に第 4 章において、高度に技術的、統計的な概念と方法が記載されているが、WG メンバーは、統計学を専門としない者がこれらの統計的な概念と方法を理解し、臨床疫学と規制科学に応用できることを目指している。この目的のために、第 3 章と第 4 章では、安全性データのメタアナリシスにおいて、適切な設計、分析、報告をするための主要な技術的・統計的な側面が記載され、続いて第 5 章ではメタアナリシスの結果を評価するための思考プロセスと得られた結果を伝える方法が記載されている。

各章の構成とその内容の概要について、第 1 章で詳細を説明する。

第1章 概要とイントロダクション

1.1 主題

2015年初めの時点で一般的な Google 検索[1]で「meta-analysis」を検索すると2千4百万件を超えるヒットがあったのに対し、PubMed[2]で科学出版物のタイトル中の「meta-analysis」を検索すると3万報を超える論文がヒットした。この主題（メタアナリシス）はそれ自体がほぼ一つの産業であると言って良いほどである。しかし、対象をヒトが服用する医薬品の「安全性」や「有害事象（AEs）」のメタアナリシスに絞ると、ずっと狭い領域となる。本書はこの狭い領域のメタアナリシスに関するものであるが、本書のコンセプトはワクチンや医療機器、動物用医薬品、さらには医薬品と医療機器を組み合わせた製品など幅広い領域で有用であることが期待される。規制当局の意思決定に資する情報提供のため、製品の安全性評価においてメタアナリシスを適切に応用する原理・原則を提示することが本書の目的である。

「メタアナリシス」という用語はそれをを用いる人によって異なる意味を持つ可能性がある。他の関連する用語には「システマティック・レビュー」や「概説」、「エビデンスの統合」がある。これらの用語は「メタアナリシス」と互換的に用いられる場合があるが、著者らはこの領域で研究を行っている多くの専門家と同様、エビデンスの統合における統計学的側面にのみこの「メタアナリシス」という用語を用いる。

CIOMS X ワーキンググループ（WG）は、「メタアナリシス」を以下のように定義することを選択した：

データが複数の独立した研究から得られたものであることを適切に考慮した分析手法を用いて、共通のリサーチクエスションに取り組む複数の研究における定量的エビデンスの統計的な結合（「用語集」を参照）。

多数の別の定義が以前の CIOMS WG を含む他の人々によって用いられているが、本 WG は、我々の興味の範囲をカバーする、シンプルで簡潔な定義として上記の定義に合意した。

この定義は次の2つの主な理由により採択された：

- メタアナリシスの概念を文献のシステマティック・レビューに続く作業のみに限定してはならない。この作業では慎重な選択基準と関連するすべてのデータの組み入れが

必要となる。本書が提示する原則は、ある単一の組織がある医薬品に関するすべてのデータを所有しているような状況において適用可能となる。

- ▶ 一般に、メタアナリシスでは元の研究による研究内比較を保持すべきであり、「粗プーリング」(治療群のすべての結果をまとめた後に比較群のすべての結果をまとめ、これらの要約された結果を比較する)とは異なる。データが複数の研究から得られたものであるという事実を無視しない限り、一般化線形モデルやベイズ法など、多数の方法が適する。

メタアナリシスはある医薬品の販売承認の前と後のあらゆる時点で実施する可能性がある。例えば、メタアナリシスの手法は、規制当局への申請時点でコモン・テクニカル・ドキュメント (CTD) に含まれる統合概括安全性情報 (ISS) に一般的に利用されるべきである。医薬品の市販後には、特に定期的安全性最新報告書 (PSUR) (定期的ベネフィット・リスク評価報告[PBRERs]とも呼ばれる) などの定期的に義務付けられた規制当局によるレビューがあるが、これらの際には、必須とされていない場合であっても、報告期間中に新たな研究が実施された時に治療後のアウトカムに関するデータを要約するのにメタアナリシスが役立つ可能性がある[3] (本 ICH ガイドライン E2C (R2) の 3.9 節を参照)。

多くの臨床的に重要な問題に対し、文献を検索し関連するすべての研究結果を見出すことは困難な作業であるが、正確な要約結果を得るためには研究の網羅性が最も重要な要素となる。例えば、ある画期的な製品の開発段階や登録プロセスにおける個別の参加者データを含むすべてのデータが入手できるならば、この網羅性の課題は低減する可能性がある。しかし、選択基準は極めて重要であり、これらがエビデンスの強さと質、すなわちメタアナリシスから得られる結論の説得力に影響を及ぼす。

本書に示すメタアナリシスの実施原則は、製品ライフサイクルのあらゆる段階において、薬事規制下の医薬品開発プログラムに適用されることを意図している。我々は多くの場合、単一の (主要な) 課題に焦点を当てて考察するが、医薬品開発では、安全性に関するシグナル検出を含む多数の目標があり、必然的に複数の時点における有害事象 (AEs) に関する多数の研究への対処が必要となる。メタアナリシスの手法を粗プーリングとしてではなく、データの評価に用いることが重要である。

1.2 概要と範囲

メタアナリシスについて現行の実務や一連の原則を記述した多数の出版物 (例[4-7]) があるものの、コクラン共同計画のものを含めて、十分なものとは言えない。なぜなら、これ

までの出版物は、特に医薬品の安全性における規制当局の意思決定に影響を与えるエビデンスの基準やメタアナリシスの説得性の問題に特化したものではなかったからである。本書はエビデンスの統合のための統計学や数学の一般的な手法について大まかに説明するが、この領域の包括的な教科書を目指すものではない。安全性データは複数の臨床試験からさまざまな方法で結合することができる。以下に、本書の原則が適応可能ないくつかの状況を例示する：

- 製品の販売承認時点での総合解析における臨床開発プログラムから観察された事象の要約。メタアナリシスの原則は、(粗プーリングとは対照的に) 有害事象と医薬品の因果関係の可能性を評価するために治療群間の比較への利用が一般に推奨される。
- 特定の有害事象を評価するための前向きメタアナリシス。一例としては、糖尿病のための医薬品開発における心血管リスクの評価がある[8]。
- 規制当局またはスポンサー（治験依頼者）は、他の治験責任医師から指摘された安全性シグナルに応じて独自のメタアナリシスを開始する場合がある。これは規制当局がいくつかの治験依頼者からのデータを利用可能な場合に特に有用となる可能性がある[9-12]。
- 他の治験責任医師はさまざまな理由でメタアナリシスを開始する場合がある（例 安全性シグナルへの対応）[13-15]。
- 安全性エンドポイントの評価のための観察研究の要約。一例としては、心筋梗塞と個別の非ステロイド性抗炎症薬の関連の可能性を評価した18集団における25の文献からの観察データを対象としたメタアナリシスがある[16]。

以下のパラグラフで、第2章以降の内容を要約する。

第2章の主たる目的は、医薬品開発および規制当局の意思決定におけるメタアナリシス（およびシステマティック・レビュー）の利用の背景と状況を示すことである。第2章では、これらの進化しつつあるアプローチによって、医薬品安全性への理解を深め、全体的なベネフィット・リスクの評価に貢献できる一方で、強度やデータソース、方法論的要件の評価にあたっては慎重に行わなければならないことを結論としている。これらすべての項目が最終的に規制当局の意思決定プロセスに影響を及ぼす可能性がある。

第3章の目的は、有益かつ有用な結果を生み出すような安全性データのメタアナリシスを

設計する方法を記述することである。メタアナリシスのプロトコルを前向きに計画することの重要性に重きが置かれている。その意図は、厳格な規範的ルールを示すことによって利害関係者（ステークホルダー）を束縛するためではなく、状況に応じた柔軟性を持つデザインのための実際的な原則を述べることにある。取り上げたトピックは、対象母集団の定義、研究の選択、アウトカムの定義とバリデーション、各研究のバイアスリスクの検討および個別参加者レベルのデータと要約レベルのデータを用いることの意味合いとその比較などである。本書で特に重視しているのはランダム化比較試験の解析であるが、場合によっては、観察研究から最も重要な（少なくとも追加的な）エビデンスが得られることもあり、第3章にはこの観察研究のメタアナリシスに特有の問題に関するセクションを設けている。また、ネットワークメタアナリシスの概念を紹介し、このトピックに関連したいくつかの重要な原則を考察するセクションを設けている。この領域には大きな関心が寄せられているが、ワーキンググループの見解では、このトピックはまだ十分に成熟していないため、ネットワークメタアナリシスの実施および活用方法に関する包括的な総意に基づくガイダンスを示すことはできない。

第4章の目的は、安全性データのメタアナリシスに伴う分析や報告に関する指針を提示することである。本章は、分析デザインが最終決定され、データが収集された後におけるデータ分析の技術的な詳細を記述する。4.10.2項には、メタアナリシスの報告書に記載すべき項目を概説した報告チェックリストを記載する。広範囲に及ぶ技術的側面の検討は、本章には含まれていないので、関心のある読者は他書を参照されたい。但し、実用的な意義や分析手法選択の重要性については記載している。第4章はどちらかと言えば技術に関する章であるが、解釈的な内容も示されている。

第5章は、専門家ではない読者を対象として、メタアナリシスの結果を評価する思考過程を説明する。推奨するアプローチは、第一印象から始まり、分析手法の徹底的な評価、他のエビデンスとの調和、得られた結果から規制措置の必要性が示唆されるかへと続く。第5章の要点は、詳細分析の前に結果の全体的な重要性を評価すること、適切な専門家をレビューに参加させること、当該製品のベネフィットとリスクに関して入手可能な他の情報とメタアナリシスの結果を関連付けることである。重要事項として、メタアナリシスの価値（説得力）に影響を及ぼす可能性がある諸バイアス、すなわちソースデータ、方法、結果の解釈におけるバイアス、報告バイアスや出版バイアスについて説明する。最後に、公衆衛生への影響がどのように認識され、結論の強さが緊急性やコミュニケーションの媒体にどのような影響を与えるかについて説明する。

第6章では、有意義なメタアナリシスを実施するために必要とされるリソースに焦点を当てる。メタアナリシスは複雑で多様な技術的専門知識を必要とすることから、ワーキング

グループは極めて初期の計画段階から適切な専門家が関与することを推奨している。チームの意思決定の指針とするための一般的な参考資料が提示されているが、具体的な問題により必要とされる専門知識が決まることから柔軟な対応が推奨される。科学的計画に加え、臨床試験実施に関する基準（GCP）の文書作成ガイドラインを順守するために専門チームのリポジトリ（文書データ保存庫）の設置が推奨される。適切な検索戦略とデータソースは成功のための重要な基準であることから、この章ではこれらに従った導入ガイダンスを行う。提案するメタアナリシスの範囲には、公開のデータソースのみならず、適切かつ入手可能な独自のデータソースも検討に含めることが推奨される。

付録 1～4 には、本書で考察されている事項を具体的に説明した実際のメタアナリシス（いくつかは本書の本文中で言及されている）の例を示す。

補遺 1 に示した用語集には本報告で使用されている重要な専門用語と定義を掲載している。但し、この用語集自体と個々の定義のいずれも包括的なものとみなすことはできない。これらは読者の理解を助ける目的で掲載したものである。

1.3 読者層

本書はやや幅広い読者層を想定している。2 つの主な層は規制当局の担当者と製薬企業の医薬品安全性担当者である。これらのステークホルダーは本報告が特に有用であると感じるはずである。本報告は医薬品安全性の医学的側面や科学的側面に経験があるものの統計や疫学の正式な訓練を受けた経験は限られている可能性がある人々を対象としている。ワーキンググループは、本報告によってこれらのステークホルダーがメタアナリシスに適用される統計的側面や疫学的側面への理解を深められることを期待している。もう一つの重要なステークホルダーグループは、統計的手法についてはよく理解しているものの影響力の強い有害事象の評価経験はやや限られている可能性のある統計学者である。明らかに、これらのステークホルダーグループ間の協力は関係する組織にとって重要であり、患者の安全性向上に有益である。少なくとも、本書には、いくつかの臨床試験のデータを組み合わせることの強みと限界を理解するための考慮事項が記載されており、質の高いメタアナリシスを行う上で、さまざまなステークホルダーグループがより明瞭なコミュニケーションをとり、協力を促進することを可能にするはずである。

また、ワーキンググループは、独立した学者や医薬品開発業務受託機関（CRO）、患者擁護団体、診療に携わる医療専門職、その他の科学者および科学ライターが本領域における研究を実施または批判的に評価する際に本報告が、彼らにとって興味深いものとなることを期待している。

1.4 将来

この領域はますます重要性を増し、複雑さは進化し続け、メタアナリシスを用いて出版された論文数は劇的に増加しつつある。データのオープン化と分析の透明化へ向けた動きは、分析に利用可能なデータ量を増やし、利用可能なデータソースを拡大させる。しかし、メタアナリシスのすべてが適切に実施され、適正に解釈されるわけではないため、メタアナリシスの拡大は課題も提起する。ワーキンググループは、本書がその意味においてタイムリーなものであることを期待するとともに、最終的に患者の安全性向上につながる質の高い研究の必要性を強調するものである。

第2章 背景

本章の目的は、医薬品の全ライフサイクルにおいて研究統合の使用のための薬事規制および科学的背景を提供することである。医薬品開発のどの時点においても、システマティック・レビューおよびメタアナリシスにより開発計画の道筋や市販後に必要とされる全ての行動指針となる重要な情報を得ることが出来る。本章ではメタアナリシスを検討すべき理由および時期に関する論理的エビデンスを示す。これは全て薬事規制における意思決定やその決定を通知するために必要な業務、データ収集と解析に関連するものである。薬事規制は本章の主要な焦点であるが、第1章で示されたように、薬事規制の範囲外でも有益な情報となることを願っている。

2.1 本書を取りまく状況

医療、バイオ・製薬業界および一般社会の意思決定者の間で、医薬品のベネフィットとリスクに関する利用可能な最高のエビデンスへの要望が増大している。通常、主要なエビデンスとなるのは、薬事規制のプロセスで得られたランダム化比較試験の結果である。全てのエビデンスを概観し、実現可能で実際的である場合、エビデンスを統合して結果を要約することが最良の戦略である。有効性においては、組み入れた各々の試験で予め決められた同一もしくは非常に類似した事象をアウトカムとして使用することが一般的である。しかしながら、安全性に関しては計画外の有害事象が不可避であるため統合プロセスは更に困難となる。

事前にアウトカムが定義されるベネフィットと比較して有害事象に関するエビデンスの統合は、元々、研究の焦点が有害事象に関するものでない場合、より困難となる。また、すべての原因による死亡以外の有害事象のアウトカムは、複数のおそらく不正確な定義という点から、様々な問題が生じる。統合されたエビデンスは、類似の他の医薬品や別の試験に関する情報といった、他の関連情報と照らし合わせることにより、特定の有害事象が医薬品曝露に関連する可能性があるかどうかを決定する一助となる可能性がある。特定の製品の有害性の大きさを評価することは極めて重要であるが、ベネフィット・リスクバランスの全面的な評価については本書の守備範囲を超えている。

システマティック・レビューとメタアナリシスに関するいくつかのガイドラインは EQUATOR のサイトから閲覧可能である[17]。これらにはメタアナリシスの報告に関する PRISMA ガイドライン[18]、その延長である個別患者データのメタアナリシス[19]およびネットワークメタアナリシス[20]の報告に関するガイドラインが含まれる。これらのガイ

ドラインは、メタアナリシスの実施よりむしろ報告の方に焦点を当てているが、計画時の検討を行う際に有益な点が多い。このサイトから多くの重要な情報が閲覧可能である。

元々の研究では答えが出なかった未解決の問題に対処するためのデータの集積と解析には、現行の医薬品規制ガイダンスは手薄である。大部分のガイダンスとコクラン共同計画のレビューはベネフィットの評価に多くの注意を向けており、有害性に関するエビデンスの統合をめぐる問題は十分に提起されてこなかったことは注目すべきである。第1章の参考文献として引用した Khan ら[7]の書籍では、発表済みのレビューを用いた有害性の批判的吟味に関する章があるが、具体的に有害性を見るこのアプローチは一般的ではない。我々が有害事象を重視するのは、現行の CIOMS レポートの貢献が大きいといえる。

2.1.1 現在の研究統合の背景

「研究統合」の歴史[21]には、物理学者の Rayleigh (1885 年) の“「新しい素材を受け入れる」だけでなく「古いものを消化吸収する」必要性”を強調する言葉が引用されている。以前から様々な研究からデータを利用する統計手法は用いられていたが、「メタアナリシス」という言葉は 1976 年に導入された[22]。1976 年時点では社会科学が先行していたが、特にランダム化臨床試験のデータを統合するため、医薬品の研究者は急速に後れを取り戻した[23]。コクラン共同計画 (www.cochrane.org) は、医療におけるシステマティック・レビューの必要性への注意喚起と実際にレビューを実行し更新し続けることにおいて、極めて重要で中心的な役割を担ってきた。コクラン共同研究は、潜在的に偏った利害関係者から独立しているという点で強い立場に立っている。

2001 年にコクラン共同計画が設立した介入による有害作用の評価を行うための小規模の非公式なサブグループ[24]は特に関連がある。この the Adverse Effects Methods Group (AEMG) という名称のサブグループは 2007 年に正式に認可され、有害事象の同定と系統的評価のための最良の実施手法に関するガイダンスの貴重な情報源となっている。AEMG は、医療におけるあらゆる介入は少なくともある程度の有害作用に関連すると推測する。有害事象の発現率が低い場合には、対照よりも本当に発現率が高いか、あるいは低いかを明らかにすることが困難となり、有害事象が介入（医薬品）によるものかどうかを決定する統計的検出力は不十分となる。これは介入と有害事象の因果関係の決定が困難になることを意味する。十分な情報を得た上での意思決定に際しては、介入による有益性と有害性の両者の系統的な評価が必要である。AEMG は、とりわけ介入による有害性と関連がある方法論的不確実性が発生した時に科学的助言を提供する。また、AEMG は観察研究のデータが妥当である可能性を認識しており、また問題によっては、ランダム化比較試験由来のエビデンスのみに制限することが不必要な限定であることも認識している。

メタアナリシスに注目することでランダム誤差や標本誤差を減少し、また結果の系統の変動が明らかになる可能性がある。もし最初の試験に（認識された、もしくはされていない）バイアスがあれば、メタアナリシスによって明らかな標本変動が恐らく劇的に減少するために不適切な結論を導く可能性があるが、バイアスは試験で再現されてきた。これは明らかに観察研究の大きな問題であるが、ランダム化比較試験でさえも（例えば、追跡調査の時期が適切でないため）影響される可能性がある。また、例え組み入れた試験が全て高品質であっても、メタアナリシスの実施のそれ自体により発生するバイアスもある。大きな注目を集めているこのようなバイアスの例として、「発表バイアス」がある。しかし、バイアスの発生源として発表過程に過剰に注目することで他の発生源を軽視することにつながる可能性がある（バイアスの潜在的な発生源に関する議論は 3.8 参照）。米国医師会雑誌（JAMA）の編集者の一人と共同執筆した解説で、著者が「結果的に、JAMA としてはメタアナリシスがアウトカム、推論および解釈について因果関係の表現を用いて報告されるよりむしろ、関連性と評すべきような観察デザインを意味すると考える。」[25]と言及したことは興味深いことである。

この分野では歴史的に、主要エンドポイントに基づいていないデータに関するメタアナリシスを行うことには強い反対がある[21,26,27]。以前から、例え臨床試験報告に関する統合基準（CONSORT）による具体的なガイドライン[28]が存在しても、発表された臨床試験報告書での有害性の報告は不適切であると認識されてきた。医薬品評価において規制当局が使用する試験の状態は可能な限り高品質であるべきだが、有害性を評価するデータに注目する際に発表データに依存してしまった場合は、すぐに関連する科学的警告を意識する必要がある[29]。

上記の議論のほぼ全てはランダム化比較試験からのデータに関連している。治療への割り付けがランダムに行われた場合のデータ（このようなデータが望ましい）を用いて関連する問題に回答できることに、反対意見はほとんどない。ランダム化データが利用可能であるにも関わらず、不確実性が残っているまたは重要な問題に対するデータがない場合は、観察研究からのデータを活用してもよい。

観察データによってランダム化データが有益に保管されるという状況の良い例は、シクロオキシゲナーゼ-2（COX-2）阻害剤が心臓血管毒性のリスクを増加させるかという問題である。138 のランダム化試験からの心臓血管リスク関連 COX-2 阻害剤に関するデータを再検討した際、Kearney らは要約相対リスクを 1.42（95%信頼区間、1.13-1.78）と評価した[30]。McGettigan らの研究においては、観察研究のメタアナリシスに基づいたロフェコキシブに関して、要約相対リスク 1.31（95%信頼区間、1.18-1.46）と本質的に同様の結果

となった[31]。更に重要なことに、観察データからロフェコキシブ投与量 25mg/日以上での要約相対リスク 2.19 (95%信頼区間、1.64-2.91) という結果が得られ、試験データではこの計算を行うのにデータが不足していた。また、観察データはリスクが治療初期に増加する可能性を指摘したが、試験データは最初、この有害性までの時間的關係を捉えることがなかった。この発見は他の観察研究により、更なる確認がなされた[32]。

観察データのメタアナリシスが非常に有効である特に良い例に、ホルモン領域がある。2つの実薬治療を比較した時に経口避妊薬を併用したランダム化試験は可能であるが、プラセボと比較することは（妊娠を望まない女性がプラセボを投与されるため）一般に倫理的ではない。倫理的配慮を別にしても、ランダム化比較試験は一般にほとんど効果が望めないため不適切で、例えば乳癌のリスクがある経口避妊薬併用での効果の観察研究は大きな物議を醸した。少なくとも部分的な議論のため、オックスフォード大学の癌疫学ユニットの Valerie Beral 博士は、観察研究から個々のデータを統合する目的で多数の研究者の集団[33]を集めた。その後、経口避妊薬併用に関連する小さな絶対リスクおよび相対リスクが示された[34]。

また、Beral 博士のグループはホルモン補充療法と乳癌のリスクに目を向け、利用可能な個別データがあることの利点を示した。Beral 博士は、更年期障害後の経過時間は年齢よりも関連性がある可能性を明らかにした。この解析は原著者によって実施されていなくとも、更年期障害発現日が研究のデータに記録されていたために可能であった。観察データではなく個別患者データ（患者レベルのデータ、参加者レベルのデータもしくは個人レベルのデータとも呼ばれ、ここでは IPD と略す）がランダム化試験に対処することに関する最近の PRISMA の声明が注目される可能性がある。統計手法は、ランダム化比較試験のデータによる IPD のメタアナリシスに用いたものと同様であるが、解釈に十分な注意をすべきであることは明らかである[35-37]。

2.2 医薬品開発と規制プロセスを取りまく状況

医薬品開発中、ランダム化比較試験のデータ収集は主として効果の主張を裏付けることを目的とするが、有害事象、臨床検査値、患者報告アウトカム、バイタルサイン、心電図といった安全性関連情報も収集される。これらの情報は全て重要であるが、我々は症例報告書に有害事象として入力された情報に注目する。薬理学の知見、クラス効果その他の情報から、重要であるという確信に至った幾つかの有害事象が積極的に収集される。有害事象のコーディングは試験のプロトコルで入念に定義されたアウトカムではない場合、多少困難となる。また、多くの有害事象は事前に規定されず（できず）、結果として自由記載で入力され、その後、解析を容易にするため ICH 国際医薬用語集 (MedDRA) のような辞書

を用いてコードされる。

治験報告書によりこれらの有害事象が要約され、治験中に特定された有害性の大きさについて統計的推定値が提供される。これらデータの要約は各々の研究および開発プログラム全体に対して行われるが、多くの場合は規制目的で、正式な統計解析というよりむしろ表形式で物語風である。単純な表形式表示に基づいた解釈は、Lièvre ら[38]および粗プーリングを活用した表示が誤解を生じる可能性を示した Chuang-Stein & Beltangady の文献[39]で議論されたのと同じことが起こる可能性が高い。

安全性確保の目的において、ランダム化比較試験のデータには重大な限界がある。併存疾患、重度の疾患、肝および腎機能障害があるような患者は、一般化の可能性に影響を与えるため、通常、治験から除外される。たとえこのような患者を組み入れる場合でも、かなり少数であることから、メタアナリシスが有効であると考えられる。治験においては、治療期間や追跡調査期間、また用量が制限されるため、外挿が限定される可能性があるが、個々の研究データが利用できる場合は、メタアナリシスが有効な手段となる。医薬品の製造販売承認の前後において、規制当局、スポンサーおよび治験責任医師は、医薬品の安全性を評価するために多数の臨床試験データの要約と統合にメタアナリシスを活用することができる。

注目すべきは、メタアナリシスが薬事申請の一部として絶対的な必要条件ではないことである。例えば、日米 EU 医薬品規制調和国際会議 (ICH) E2F ガイドラインの治験安全性報告では、メタアナリシスからのデータが「利用可能となり」、今後貢献する可能性があると言及したが[40, 3.10]、メタアナリシスが常に施行されるべきとは明言されていない。同様に、(有効性評価に関してではあるが) ICH E9 では「メタアナリシスは有益である可能性がある」と述べられ、また「事前に準備されたプロトコル」の必要性について指摘している[41]。メタアナリシスは医薬品開発の一部として絶対的に必要ではないものの、ワーキンググループとしては、メタアナリシスは継続中の安全性監視の一環であるべきで、データの統合が適切な状況において薬事申請にメタアナリシスを組み入れるように助言されると考える。

安全性データのメタアナリシスは、医薬品承認申請の前後で施行可能である。医薬品や製品群に関する臨床安全性データの評価におけるメタアナリシスの役割について簡略化した図を図 2.1 に示す。製造販売前の医薬品のメタアナリシスは、関心がある有害事象のシグナル検出、因果関係解明の一助、リスク管理やリスク最小化活動の必要性、さらには医薬品のベネフィット・リスクバランスの評価のために実施される。この評価は、医薬品開発の継続あるいは中止の意思決定、申請時の申請資料、企業と規制当局間での議論のために

有用となる。

製造販売後のメタアナリシスは、通常、臨床試験からの安全性情報を更新する目的や新たな安全性の知見を調査する目的で施行される。市販後の安全性の知見は、臨床試験では調べられなかった稀な有害事象、もしくは例えば、心臓のイベントのように背景発生率が高い事象において、少し頻度が高い有害事象に関連していることが多い。これらの状況においては少しの違いを明らかにするために、多数の曝露された被験者が必要である。新規臨床試験は製造販売後において一般的で、メタアナリシスは更にベネフィットとリスクを詳細に説明するためにこれらの結果と共に施行される。同様にメタアナリシスは、死亡数に関する様々な治療の効果および明らかな急性心筋梗塞、もしくは急性心筋梗塞疑いの患者の主要疾患を調査した **International Studies of Infarct Survival (ISIS)** [42-46]と同じように、有効性と安全性の両方で追加的な試験の仮説を立てるために使用可能である。

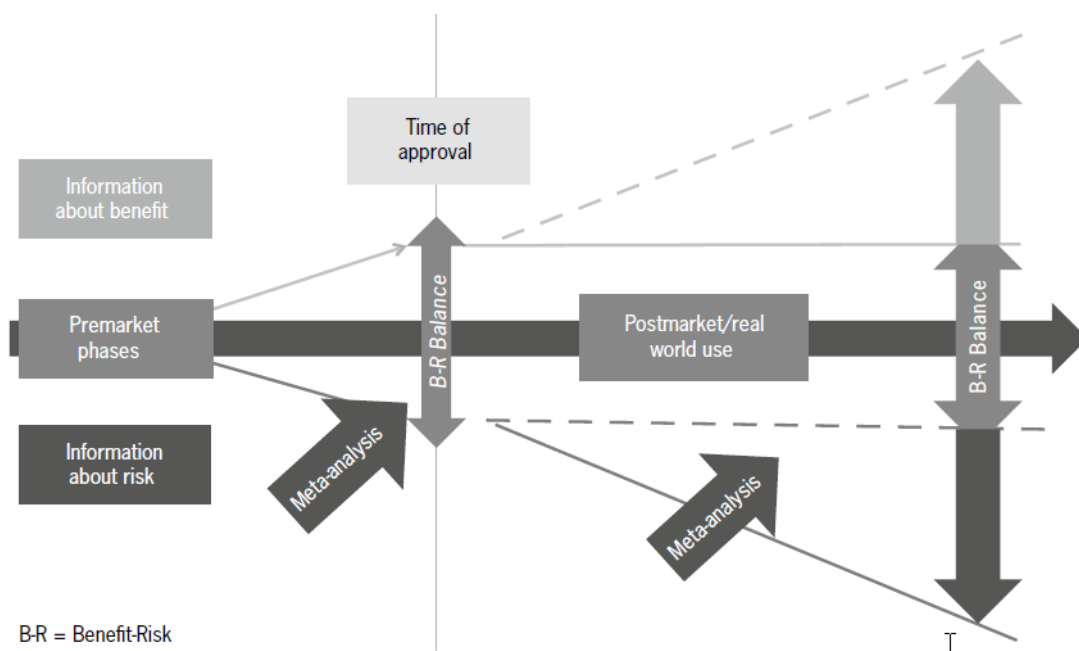


図 2.1 医薬品または薬に分類される製品に関する臨床安全性データの評価でのメタアナリシスの役割の簡略図

出典：Hammad et al. The future of population-based postmarket drug risk assessment: a regulator's perspective. Clin Pharmacol Ther, 2013 [47].

規制当局の多くは、未認可の製品に関する全ての関連試験データを所持している重要な立場にいるが、アメリカ食品医薬品局（US FDA）は例外として、メタアナリシスを施行する伝統がない。個別データを用いてメタアナリシスを施行することは可能だが、日常的な業務ではなく、データ収集および解析の施行には、恐らく大抵の場合は2~5人/年が必要となる。新薬承認申請書類の一部となっている試験結果の粗プーリング（治療群および対

照群での全事象を加算し、個々の研究結果を無視する)は、スポンサーによって申請時に施行される場合が多い。このような粗プーリングでは交絡を含む無効な結果を生み出す可能性のため、一般的に推奨されない[38,39,48(Figure1)]。恐らく再解析することは可能であるが、規制当局は通常、再解析を実施しない。

過去 10 年間で、メタアナリシスの結果から医薬品の有害事象に関連する多くの規制上の安全性問題が提起された。メタアナリシスは多くの場合、エビデンスを追加し、医薬品の安全性に関する意思決定のため情報を提供してきた。しかしながら、同じ問題に対して様々な解析を行った結果、異なる結論が出た際には、メタアナリシスは議論を生み出し、混乱さえも招くことがある(例:[10,49])。このことは、メタアナリシスを実施する際に高い水準の解析が求められることを示している。

メタアナリシスに続く規制当局の審議により、承認済み添付文書(例:アメリカの添付文書、EUの製品概要)の改訂という結果になることもあるが、全く改訂なしの場合や、その対極で製造販売の取消しということも起こり得る。

規制当局および医薬品のスポンサーは、安全性データのメタアナリシスを独自に施行するのに加えて、科学文献やその他の規制領域外の研究者や市民団体により発表されたメタアナリシスの結果に対応するように強いられつつある(本書の付録3および4は、これらが生じた場合の具体例である)。これらに対応することは、製品のベネフィット・リスクバランスに与える結果の影響だけでなくメタアナリシスの品質と頑健性を評価することを意味している。この影響評価は過去になされた規制当局の決定により構成される可能性があり、あるいは、他のエビデンス資料と併せて将来の意思決定のために情報を提供する可能性がある。

医薬品安全性データのメタアナリシスはその頻度が増えることともに、メディアでしばしば紹介されることが期待される。何故なら、ランダム化臨床試験(RCT)と市販後データに関して透明性が高まり、開放性がより拡大するからである。

規制当局の科学者、医薬品のスポンサーもしくは学術機関を含む利害関係にある組織がメタアナリシスを実施する。原則として、これら全ての組織で信頼のおけるメタアナリシスが実施可能であるが、時としてプライバシーや知的財産権のため、アクセスできるデータが組織によって異なる場合がある。データへのアクセスが組織によって異なることは、研究の組み入れに重要な影響を与える。レビュー毎に組み入れた研究が異なる場合、別の結果、ひいては別の結論をもたらす可能性がある。

発表の有無にかかわらず、また知見が何であれ、全ての関連研究の結果を入手することが重要であることは明確である。例え臨床試験データが一般的に個別データに基づいていようと要約データが依然明らかでなくても、治験責任医師であれ産業界を含むスポンサーであれ、このようなデータを利用可能にすべきという議論があり、これから数年の内には利用可能性が増加しそうである。米国医学研究所は、透明性とデータ共有に関する実行可能な枠組みを計画する論考を発表してきた[50,51]。この委員会が 2015 年 1 月に発表した報告[52]は、少なくとも過去に行われていない新しい試験である場合に関しては、多くの産業界のスポンサーが発表していないデータを公開することを表明した時期と一致する。例えば、いくつかの機関はウェブサイト（Clinicalstudydatarequest.com[53]）を介したデータ入手の共同の方法を発表し、またエール大学の科学者らは、エール大学オープンデータアクセス（YODA）プロジェクト[54]を介して業界とのデータ共有に協力する。アメリカとヨーロッパにおける業界の事業者団体（米国研究製薬工業協会および欧州製薬団体連合会）は、重要な臨床試験データの共有に関する共同原則[55,56]に記載のように透明性とデータ共有に対する全般的な取り決めに結んできた。

欧州医薬品庁はこの領域において多くの手順を踏んで、2014 年 10 月に「ヒトへの使用に対する医薬品の臨床データの出版に関する欧州医薬品庁の政策」[57,58]というプレスリリースを発表した。

個人レベルのデータを共有するための適切なプロセスは、今後もなお詳細に検討する必要がある。個人情報と他の情報源との連携は、機密漏えいに繋がることが懸念される。

今後、医薬品業界以外で多くのデータが利用可能となり、医薬の安全性を評価するためにデータ解析をしたいと考える人への手引きを本書が提供する。原則的には、個々のスポンサーは通常、既に利用可能なデータを保持している、しかし、例えば、企業合併で使われなくなったコンピュータシステムでは関連データを簡単には抽出不可能であるとか、医師主導治験でデータが利用できない可能性あるなど、実際にはさほど簡単なことではない。

CIOMS ワーキンググループ VI[59]と安全性計画・評価・報告チーム（SPERT）[60]の両者は、医薬品開発過程の早期に安全性シグナルを同定するため、スポンサーが明瞭で連携したプログラム全体の取り組みを実行することを提言した。SPERT は更に、事前定義された「安全性解析計画プログラム（PSAP）」の作成を盛り込むことを推奨した。SPERT はランダム化比較試験データのメタアナリシスによって多くの安全性の問題に最良の答えを与えることを認識し、計画への助言を行った。CIOMS ワーキンググループ VI もまた、報告書の第 6 章で医薬開発におけるメタアナリシスの使用に関して提案した。

2.3 なぜエビデンスの統合が必要か

大半の生物学的および医学研究の単一の研究で、関連する問題に対する明確な回答を与えることはほとんどない。例え研究の目的とする課題に対する回答が明確に得られたとしても、収集されたデータは他の問題に回答するために転用される。後者の利用には幾らか反論があるが、研究者および研究参加者の時間や人的資源を費やして得られたデータは大変貴重であり、将来ほとんどの場合、利用可能なデータを最大限に活用することが必要となる。それにともない、全てのデータを適正に最大限に活用するための倫理的規範について議論される可能性がある。

製造販売承認の前後で、医薬品の安全性評価に対して厳しい注意を払うことが重要である。安全性に関するエビデンスの評価では、期待されるベネフィットとの関連で医薬品の起こりうる有害性に関するデータ全体を網羅するべきであるため、エビデンスの統合が必要とされる。初期の試験で調査されず検出されなかった有害事象、または開発計画から除外された患者サブグループで発現する可能性のある有害事象などの情報の欠落を特定することも重要である。注目されつつある情報欠落の実例として、安全性の問題の容認可能な妥協点に関する患者視点や患者意向への洞察力不足が挙げられる[61]。これは、どのエンドポイントを測定するのが重要かを決定する際に、患者の関与が不足していることに由来する可能性がある。これらの情報不足により、日常の診療における医薬品使用に安全性における不確実性という結果をもたらす。上記に関連する実際に発生する可能性のある全ての問題を対象とすることは、本書の範囲外である。

2.4 研究の統合における重要な留意点

ランダム化比較試験は強力な因果推論を支持する可能性、すなわち、安全性の文脈では、医薬がイベントに関連するのではなくイベントを引き起こすことを支持する可能性がある。(観察研究において医薬がイベントに関連することは容易に起こりうる。)ランダム化比較試験での治療の割り付けは無作為であるため、患者により異なる既知未知の要素には偏りが無い。ランダム化試験においてバイアス低減のさまざまな取り組みがなされてきたが、ランダム化試験でさえもバイアスと無縁ではない。例えば、欠落データ(例:脱落者によるもの)は重要な問題に対する回答の解釈にバイアスを与える可能性がある。バイアスが起こりうる場合には、試験の組み入れの選択、関連データの抽出(およびその関連の有無の基準を決定すること)の全てが問題となる。

バイアスに関しては、次章(3.8)でより詳細を述べる。ランダム化試験におけるバイアスのリスクに関しては、コクラン共同計画により多数の取り組みがなされてきた。コクラン・

ハンドブックの第8章[6]の全てがこの問題に充てられている。

さらに、バイアスが起りうる様々な重要な領域を調べるツールがあり[62]、これはコクランレビューにおいて数多く使用された。特定の治療上の問題に関連する臨床または研究デザインの問題があるので、包括的なバイアス評価ツールは全ての状況に十分対応できない可能性がある。バイアスのチェックリストに資金援助者を含むべきかどうかをめぐる論争が続いている[63,64]。

従来のメタアナリシスの方法は、単一の要約効果量（エフェクトサイズ）を得るために結果を統合しようと試みるものである。患者によりその特性が異なり、治療に対する反応が患者により異なる可能性があるため、試験の結果がどの程度一致しているかを検討することが重要である。この状況において、単一の要約数では不十分である可能性がある。異質性に関する情報およびその調査法に関しては4.8参照。

原則として、ランダム化比較試験には試験のデザイン、実行、データ収集、解析および報告の全ての側面を定義する明確に規定されたプロトコルがある。事後合理化の可能性および報告するデータの選択肢が限られているため、主要アウトカム（いくつかの試験では複数個の主要アウトカムが存在）は事前に規定される。実際には、承認申請に使用された試験はプロトコルを厳しく遵守する傾向にあるが、必ずしも発表された試験で同じではない。これには、プロトコルが提出され、査読済科学雑誌にアクセプトされるような高品質と考えられるものであっても、最終報告書ということになるとプロトコルを順守しない可能性があるというエビデンスがある[65]。プロトコルと最終報告の間の矛盾と同じような問題は、メタアナリシスによっても起りうる[66]。あらゆる種類の試験でプロトコルを変更する正当な理由がある可能性があるが、全ての変更点は文書化されるべきである。そのような矛盾が存在するという事実は、最初の試験およびそのメタアナリシスの両方でバイアスが生じる可能性があることを意味する。それ故、単純な解釈は避けなければならない。

発表されたランダム化試験に関して、試験の各治療群における事象有と無での抽出数は理論上シンプルであるべきである。2値データは、安全性の問題に関して最も頻繁に使用される種類のデータである。これらのデータは公表物には挙げられていない可能性があるが、規制当局とスポンサーの両方でデータを利用できるべきである。いくつかの公表物において、1人の患者が何度も事象を再燃している場合、複数の独立した事象として報告されることがある。しかし、メタアナリシスを実施する際、例外を除き、事象数ではなく、（少なくとも1個の）事象がある1人の患者を数えていることを保証するためのデータ抽出に注意が払われなければならない。コクランレビューグループには大きなチームがあり、一つのレビューに数カ月間をかける。このことは、メタアナリシスを正確に行うことは簡単で

はないことを示唆している。与えられた表から要約統計量を再計算するのは実に簡単であるが、表に確実に正確な数字が記載されていることを確認するには多大な努力を必要とする。観察研究において、関連データの抽出（および確認）はかなりの困難を伴う。

2.5 結論

安全性データのシステマティック・レビュー及びメタアナリシスは、多くの医薬品のライフサイクルにおいて周知されるようになってきた。実際、これらの取り組みは医薬品の安全性プロファイルの理解を高める非常に洗練された方法に発展し、全般的なベネフィット・リスク評価に大きく貢献することが可能となっている。しかしながら、メタアナリシスは異なる強さ、異なるデータの情報源、異なる方法論的必要性、異なる意思決定プロセスへの影響を持つ可能性がある。したがって、患者の安全性確保を最大限とするため、メタアナリシスにおけるデザイン、データ選択、方法論的アプローチ、結果解釈の基準など、あらゆる側面に注意を払うべきである。

第3章 計画

本章は、2つの異なるタイプのメタアナリシス（後ろ向きおよび前向きメタアナリシス）についての簡単な説明から始める。続いて、メタアナリシスの準備の中で完全なプロトコルを定義するための要素リストを示す。次に、薬事規制下の安全性評価に関連するデザインの側面に焦点を当てる。

- 3.1 メタアナリシスのための準備
- 3.2 対象母集団の定義
- 3.3 アウトカムの定義
- 3.4 期間の異なる試験
- 3.5 研究の選択とデータの抽出
- 3.6 研究の規模と小規模研究の効果
- 3.7 多群比較試験
- 3.8 メタアナリシス、個々の研究、研究全体のバイアスリスク
- 3.9 個別参加者データ、要約レベルデータまたはその両者へのアクセス
- 3.10 観察研究のメタアナリシスに特有の問題
- 3.11 ネットワークメタアナリシス

3.1 メタアナリシスのための準備

主要事項の要約

- ▶ 薬事規制下における医薬品安全性評価のためのメタアナリシスには特有の問題がある。例えば、メタアナリシスを実施するにあたり事前の研究計画の立案、メタアナリシスにおけるランダム化比較試験（RCTs）の個別参加者データへのアクセスおよび探索的メタアナリシスと検証的メタアナリシスとの区別である。
- ▶ 後ろ向きメタアナリシスから前向きメタアナリシスの異なるタイプのメタアナリシスがある。前向きメタアナリシスは、後ろ向きメタアナリシスの認識された問題のいくつかを克服することができ、多くの利点がある。
- ▶ メタアナリシスのタイプに関わらず、綿密にかつ事前に計画されたプロトコルは、メタアナリシスの実施のために常に必要である。詳細な統計解析計画（SAP）をプロトコルとともに作成しなければならない。
- ▶ メタアナリシスのデザインにおけるいくつかの主要な要素を、メタアナリシスのプロトコルに含めなければならない。

- ▶ 利用可能な試験を知る前にシステマティック・レビューのためのプロトコルを公表することは、レビュー著者によるバイアスを減少させ、透明性を促進するため重要である。最近、システマティック・レビューのプロトコルの国際登録が、英国保健省国立保健研究所（National Health System's National Institute for Health Research）と共同で、英国 York 大学の Centre for Reviews and Dissemination に設立された[67]。メタアナリシスが実施された後、公表される予定がある場合は、プロトコルを登録しなければならない。
- ▶ メタアナリシスは公表されたプロトコルを遵守しなければならないものの、プロトコルの変更は時々必要となる。プロトコルの変更は、その変更が研究の調査結果にどのように影響するかに基づいて行われなければならないことが重要である。メタアナリシスのプロトコルの変更は全て、完全に文書化され、各変更の合理性を示さなければならない。

メタアナリシスの文献は数多くあるが、コクラン共同計画から公表されているような、現在の手順および諸原則では十分ではない。なぜなら、特に医薬品の安全性に関する規制の意思決定のために、メタアナリシスの根拠基準および説得力に関連する問題に具体的に取り組んでいないからである。本章では、メタアナリシスを実施する前の最初のプロセスであり、メタアナリシスのプロトコルの一部であるべきメタアナリシスのデザインの主要要素を記載する。目的は、既存の確立したメタアナリシスの詳細を繰り返すことではなく、むしろ薬事規制下での医薬品の安全性評価のためのメタアナリシス特有の問題を強調することにある。例えば、プログラムレベルのメタアナリシスの前向き計画、RCTs のメタアナリシスのための個別参加者データへのアクセス、および探索的メタアナリストと検証的メタアナリシスとの区別である。

メタアナリシスはしばしば、そのゴールを考慮して計画されていない研究に対して、後ろ向きに実施される（本書のゴールの一つは、前向きの方向にメタアナリシスを向かわせることである）。加えて、多くのメタアナリシスは、公表された文献から抽出した要約統計量を基としている。本書では、既に結果が既知である研究について計画、実施されたメタアナリシスを後ろ向きメタアナリシスと表記する。前向きメタアナリシスは、結果が判明する前に、メタアナリシスの適格性が確認、評価、決定された研究（一般に RCT）のメタアナリシスである[6]。

前向きメタアナリシスは、試験の結果が判明する前に、仮説を前もって具体化し、し、試験選択基準の事前適用、および試験結果が明らかとなる前に解析計画を指定できるようにすることによって、後ろ向きメタアナリシスで認識された問題のいくつかを克服することができる。これにより、潜在的なバイアスのかかった、特定のサブグループもしくは特定のエンドポイントによる強調を回避することができる。前向きメタアナリシスアプローチ

は、他の利点もある。重要な利点は、前向きで正確なアウトカムデータの収集の可能性である。加えて、多くの前向きメタアナリシスは、個別患者データ（IPD）を解析することである。要約レベルとは対照的に、患者レベルデータの使用は、患者レベルでの異質性の調査およびイベントまでの期間の解析（時間事象解析）が可能となるなど多くの利点がある（詳細は 3.9 参照）。前向きメタアナリシスは、個々の試験では対処することができないリサーチクエスチョンに対する十分な検出力を持つ計画（例えば、まれな臨床アウトカム、安全性エンドポイント、サブグループ解析の計画）を立てることができる。結果に関わらず、解析を公表する計画を持つことは、出版バイアスを回避するための主要なステップである。また、各試験におけるデータ収集の標準化は、前向きメタアナリシスにより促進される[60]。

前向きメタアナリシスの特別な場合は、構成要素である試験自身がメタアナリシスに使用されるよう計画されていて、対象とするアウトカムに対するバイアスのない治療効果推定をデザインすることができる場合である。参考例は、US FDA ドラフトガイダンス、**Diabetes mellitus: Developing drugs and therapeutic biologics for treatment and preventions (2008)[8]**に記載されている。

実際には、後ろ向きから前向きまで様々な(spectrum)のメタアナリシスが実施されている。メタアナリシスに、事前に結果が判明している試験群と事前に結果が判明していない試験群がある場合には、各々の解析と全体の解析を実施し、両者の結果を含むことが望ましい。ある研究では個別患者レベルのデータ（IPD）を有し、他の研究では要約レベルのデータのみを有することがある。これらを統合する統計手法が開発されている[68, 69]。

複数の研究に対する研究として、メタアナリシスは他の研究活動と同様の研究デザインの原則に導かれることが論理的である。メタアナリシスの種類に関わらず、メタアナリシスを実施する際にはガイドとなるプロトコルが常に必要である。綿密なプロトコルは、予測の明確化、透明性、厳密性、妥当性の向上に役立ち、メタアナリシスの一貫したアプローチを促進する。全てのデータが存在する場合でも、最低限、メタアナリシスの主要部分を定義するプロトコルが必要となる（例えば、研究種類、選択基準、除外基準、統計解析計画など）。

特に後ろ向きメタアナリシスにおいて、適格可能性ある研究の結果を事前に知ることは、例えば、システマティック・レビューの論点定義、それに続く試験の適格性基準、解析の介入比較の選択などに影響し、レビューにおいて報告するアウトカムにも影響する可能性がある。このようなバイアス発生の可能性があるため、研究者は事前に全ての方法論を確定し、文書化することが重要である。加えて、利用可能な研究を知る前にレビューのため

のプロトコルを公表することは、レビュー著者バイアスの影響を減少させ、方法および手順の透明化を促進し、重複の可能性を減少させ、計画された方法について査読評価を可能とする[70]。システマティック・レビューのプロトコル国際登録（PROSPERO と称される）は、英国 York 大学の the Centre for Reviews and Dissemination に 2011 年に設立された[67]。

メタアナリシスのプロトコルの詳細な記載事項は、多くの公表物から入手できる（例えば、the Cochrane Handbook など）[6]。完全性を期すため、ここではハイレベルの説明を行い、続いて薬事規制における安全性評価に関連するデザインの特定の側面に焦点を当てる。

メタアナリシスのプロトコルには、以下の要素を含むべきであるが、これらに限定されない。

- **背景：**この項ではメタアナリシス実施の薬事規制下における背景と理由を記載する。作用機序、適応症、用量など研究治療の概要を示す。また本研究で解明する安全性の問題に関する既知のエビデンスについて簡潔に述べ、結果がすでに明らかとなっている試験についてはその点を明確に記載する。
- **目的：**この章は、メタアナリシスの範囲と性質を定義する目的を述べる。主目的の正確な記述から開始し、一つ以上の副次目的を述べる。リサーチクエスションでは、集団の種類（参加者）、介入（および比較）の種類、対象のアウトカムの種類を特定する。PICO（Participants、Interventions、Comparisons、Outcomes）は、これらを思い出すのに役に立つ。ある著者は、PICO を PICOS もしくは PICOTS に拡大した。T は Timing、S は Study Design もしくは Setting の頭文字である。
- **適格性基準：**適格性基準は、リサーチクエスションの側面とリサーチクエスションに対応する試験種類の仕様の組み合わせである。これは、研究集団、治療（より広義には介入）、比較、アウトカムおよび試験の種類を含む。
- **対象となるアウトカム（以下、アウトカム）の定義、調和、及び判定：**本項では、アウトカムを定義する。被験者、治療（用量）、アウトカムおよびフォローアップ期間の併合可能性の基準を詳細に示し、アウトカムの評価手順（例えば、治療薬群の盲検化）を記載し、アウトカムを評価するために設計された試験と試験終了後にアウトカムが評価された試験を区別する。複合アウトカムについては、その使用および個々の要素使用の正当性を慎重に考慮するとともに注意深く扱わなければならない。アウトカムの定義に関連して、症例の確認におけるバイアスを最小限にするため、プロトコルもしくは代替文書にて、臨床専門委員会の憲章（the charter チャーター）と判定の手順を定める必要がある。
- **データの情報源および試験選択：**この項では、メタアナリシスに含める全ての関連す

る研究（公表および未公表）および、選択された研究からの関連データの抽出手順（計画されたアブストラクターの数を含む）を定める。

- **個々の研究におけるバイアスの可能性の評価：**この項では、研究デザイン、および潜在的バイアスの評価方法を記載する。
- **統計解析計画：**詳細な統計解析計画（SAP）はプロトコルとともに作成すべきである（注：詳細な解析方法については、第4章を参照）。SAPは以下を記載すること。
 - 主要解析集団（例えば、**intention-to-treat**、**as-treated** もしくは **per protocol**）。安全性のクエスチョン論点に対する“**per protocol**”集団の使用もしくは“**as-treated**”解析が良いかについて議論がある。これらの集団はすべて、安全性の評価においてバイアスとなる可能性がある[59、71-76]。
 - 希望する集団のデータが入手できない場合の対応。主解析に加えて、重要な感度分析のための集団を特定すべきである。感度分析（もし主解析でない場合）は、最小の選択決定に対応しなければならない。言い換えれば、元の試験からできるだけ多くのデータを使用し、可能な限りメタアナリストの意見を少なくすることである。それによって初めてこれらの意見がどのような影響を持つか明らかになる。
 - 変数の運用上の定義と適切なコーディング辞書
 - 各試験のフォローアップ期間のバランスと比較群間の関連する交絡の調査計画
 - 個々もしくは要約レベルでのデータ欠落の取り扱い
 - 要約効果指標（例えば、リスク差、オッズ比、リスク比）の選択
 - 統計学的手法の詳細
 - 統計モデルの重要な仮説の評価計画
 - 実施する全てのサブグループ解析および各試験の比較群サブグループ間におけるリスク因子分布のバランスを調査する計画
 - (任意)指定された範囲のリスク増加を除外する検出力。これは、検出可能な別のリスク推定量にサンプルサイズがどのように関連するかを示す値のとり得る範囲をカバーする検出力のプロットを示すことができれば役に立つだろう [77]
 - 寡少データ数もしくはイベントが無い試験の取り扱い
 - ベイズ流アプローチを使用する場合、仮説、予め選んだ項目とそれに関する感度分析、実施方法、結果報告の詳細を提示すること
 - 感度分析の計画
- **臨床および統計的異質性：**この項では、統計的および臨床的異質性を評価するアプローチを記載し、選択した統計手法（固定効果；**fixed-effect** あるいは変量効果；**random-effects** モデル）の特性を議論し、検討する試験および対象患者レベルの共変量を（例えば、メタ回帰を介して）決定する。

- ▶ **結果の提示と報告の計画**：この計画には、例えば、各試験の要約データ、全体およびグラフ表示した（forest plots）要約データ含む場合がある。点推定値および区間推定を含めなければならない。公表計画も詳細に記載すべきである。

メタアナリシスのプロトコルは一般に登録されるべきである。メタアナリシスもしくはシステマティック・レビューでは、公表されたプロトコルを遵守すべきであるが、プロトコルの変更が時々必要となる。これは、参加者登録、データ収集もしくは未知イベントの割合の問題など、予期せぬ状況に対応するために時々変更する必要がある無作為化試験のプロトコルの場合に類似している。事前に決定したプロトコルを遵守するために、あらゆる努力が払われるべきであるが、これは常に可能もしくは適切であるわけではない。プロトコルの変更は、その変更が研究結果にどのように影響するかを基礎にしないことが重要である。研究結果の影響力が明らかとなった後に実施される事後決定、例えば、メタアナリシスもしくはシステマティック・レビューから選択した試験群を除外することは、研究結果にバイアスの影響を及ぼしやすく、避けるべきである。メタアナリシスのプロトコルの変更は、臨床試験プロトコルの修正（amendments）と同様、完全に文書化すべきであり、各変更の根拠は、公開された報告書を含め、完全な透明性を確保する形で提供すべきである。

プロトコル要素を高レベルで定義した後、この章の残りの部分では、薬事規制下の安全性に関するメタアナリシスの中で特に困難な要素のサブセットに焦点を当てる。

3.2 対象母集団の定義

主要事項の要約

- ▶ バイアスの導入を避けるために、研究戦略を特定する前に対象とする母集団を定義しなければならない。

研究戦略を特定する前に、対象とする参加者集団の定義が必要である。このステップは、バイアスの導入を避け、研究結果（所見）に基づく決定が適切な集団に適用される可能性を高めるために極めて重要である。プロトコルには、対象とするアウトカムに関連する可能性のある人口動態の詳細および重要なベースライン属性を含め、主解析および主な副次解析に含まれる対象集団セットを記載しなければならない。例えば、解析をハイリスク集団に焦点を置きたい場合、プロトコルで、“ハイリスク”がどのように定義されるかを明記しなければならない。

どの集団（およびどの試験）を評価に含めるべきであるかの基準は、コクラン・ハンドブック（Cochrane Handbook）[6]に詳細に記載されている。一般に、この基準は臨床現場で治療を受ける様々な患者を包含するために、十分に広くなければならないが、試験を総体として考えた場合、意味のある回答が確実に得られるように十分に狭くなければならない。重要なステップは、広範囲な対象集団と設定（setting）を定義することである。これは、特定のグループが対象母集団であるかどうかを決定することを含み、年齢、性別、人種、治療疾患（例えば、“重症”疾患、あるいは、狭心症、息切れなどの特定症状の存在）などの要因に基づいている可能性がある。対象は、地域社会、病院、養護施設、長期療養施設もしくは外来患者などの特定の設定（setting）に焦点が置かれる。“参加者タイプ”の基準を作成する際に考慮すべきいくつかの要因を Cochrane Handbook、Box 5.2a[6]を引用して Box3.1 に示す。

対象とする参加者のタイプは通常、各々の試験の参加者に関連する適格性基準で直接決定される。しかし、対象とする集団に部分的にしか合致していない研究に対処するルールを事前に指定することは困難である。例えば、16 歳以下の小児にフォーカスする場合、16 歳上限を事前に指定しても、12 歳から 18 歳を対象とする研究を取扱う戦略を決定できない。この場合は、“参加者の大部分は 16 歳以下である”などというフレーズで十分である。一度解析をした後に組み入れの決定をすればレビュー著者バイアスとなるが、適格性の決定では、レビューの目的に適う常識的な戦略が恣意的なルールを上回る可能性がある。困難な諸決定について、レビューで文書化すべきであり、レビューの結果（所見）に対してこれらの決定が及ぼす影響の程度を、感度分析により評価することができる。（もちろん、最善の解決策は、試験(治験)の個別参加者データ（IPD）にアクセスし、対象とする特定の参加者データを選択することをメタアナリストに許可することである）[6]。

具体的な集団の特性（characteristics）や設定(setting)に対する制限は、妥当な根拠に基づくべきであり、その根拠は、プロトコルに明記すべきである。様々なサブグループ集団間の影響に重要な違いがあるかどうか不明な場合、メタアナリシスにおいて、関連するすべてのサブグループを含めることが最善であり、その上で、効果の重要でもっともらしい差異を検定する（4.8 サブグループ解析を参照）。これは、事前に計画され副次目的として記載すべきである[6]。

Box 3.11 対象母集団の基準を作成する際に考慮すべき要素

- 疾患/状態の定義方法
- 対象集団（参加者）を記述する上でもっとも重要な特徴は何か
- 関連する人口統計学的要素（年齢、性別、民族性）はあるか
- どのような設定（setting）か（入院、地域社会で生活等）
- 誰が診断すべきか
- レビューから除外すべき患者のタイプ（介入に対して異なる反応を示す患者集団等）は他にあるか
- 関連する参加者のサブセットのみを含む研究をどのように扱うか

出典：Box 5.2a of the Cochrane Handbook [6].

上述した要素は、（医薬品を使用すべきでない患者集団を特定する）など、医薬品を使用する患者集団へのメタアナリシスの結果の一般化可能性に影響を及ぼす。

3.3 アウトカムの定義

主要事項の要約

- 重要な安全性アウトカムを適切に測定する明確な仕様が求められる。
- 生物学的な意味を持つ場合に複合アウトカムを用いるべきであり、プロトコルにそのアウトカムを選択した正当性の詳細を含めなければならない。
- 主に医薬品の有効性を研究するためにデザインされているランダム化比較試験（RCT）は、安全性評価のためにデザインされることは稀で、有害事象の収集に関する事前に特定された体系的な方法はしばしば欠落する（例えば、いくつかの臨床概念が別々にとらえられ、コード化されるなど）。事後（post hoc）メタアナリシスの評価における最初のステップは、対象とする安全性アウトカムのための、判定、検証及び確認努力があるかどうかを調査することである。

本項では、前向きメタアナリシスおよび後ろ向きメタアナリシスのアウトカム定義において考慮すべき具体的事項を議論する。まず、前向きメタアナリシスの議論から開始する。なぜなら、後ろ向きメタアナリシスにおいても、可能な限り前向きメタアナリシスと同様の基準に準拠すべきだからである[6]。

3.3.1 前向きメタアナリシスのアウトカム定義

前向きメタアナリシスは、医薬品の開発プログラム中、もしくは、いくつかの試験を共同するコンソーシアムによって市販後に開始することができる。前向きメタアナリシスでは、プロトコルの作成や研究の実施段階においてアウトカム定義に関連する問題を検討することが可能である。これらの問題には、主にアウトカムの選択と検証、データの収集方法、必要に応じて、判定方法などが関連する。

歴史的には、用語の幅広い検索（またはアウトカムの広範な定義）によって、すべての関連用語が安全性のアウトカム定義に確実に含まれるようにした。しかし、広範囲の複合アウトカムの使用は、真の相対リスクを過小評価し、各治療群での発生頻度が同等と誤分類し、安全性シグナル/問題の潜在的なマスキングを引き起こす[60、78]。臨床的に疾患スペクトルとして示されるイベントに特に関心がある場合を仮定する。狭義の分類では、疾患スペクトルの末期で、より重篤なイベントのみを含むが、広義の分類では、臨床スペクトルの全体を含む可能性がある。広義の分類では、本来の（しかしおそらく未知の）作用メカニズムに関連性が低いイベント、あるいは、単にその性質上、治験で誤分類される可能性が高いイベントを含める可能性がある。この状況では、狭義の分類に比べて、広義の分類に従って観察された相対リスクは1に近似する。興味深い実例として、「The Statistical methods for monitoring clinical trials」[79]の9章を参照のこと。

アウトカムの選択と検証

一般的に、重要な安全性アウトカムの適切な測定に関する明確な仕様が求められる（例えば、心筋梗塞の場合、心筋の酵素を測定することが重要となる可能性がある）。利用可能な場合は、アウトカムには標準的な医学的定義を使用すべきである。医学文献で十分に確立された定義がない場合は、標準化されていない定義の使用について、関連する規制当局および他の利害関係者の合意を得ることが重要である可能性がある。また、選択した定義の臨床的関連性を評価するために検証作業を実施すべきである。治験中に収集されなかった関連情報の不足のために、特定の診断アルゴリズムの性能を評価する能力が操作上の課題によって妨げられる可能性があることは注目に値する。

代替アウトカムの使用

臨床アウトカムの十分な情報が入手可能な場合、代替アウトカム（例えば、潜在的肝障害の代理として肝酵素の上昇）は推奨されない。代替アウトカムは、意図するアウトカムから一歩離れるため、別の変動性が加わる。さらに（可能であったとしても）妥当性の検証は困難になる。

複合アウトカムの使用

一般的に、複合アウトカムは、注意して使用すべきである。使用する場合は、生物学的に

意味を持つべきであり、プロトコルに複合アウトカム選択の正当性を十分に記載すべきである。複合アウトカムを使用する場合、個々の要素の解析も計画に含めるべきである。総合結果が要素の一つによって決定される、もしくはいくつかの要素が他とは異なる方向に向かう状況を特定することが特に重要である。実際には、ある要素のリスクが増加し、別の要素のリスクが低下する場合、複合アウトカムの総合評価では、相関は見られないとの結果となる可能性がある。複合アウトカムの定義が、対象となる特定の有害事象の解析の結果、感度および特異度にどのように影響するか、つまり薬剤と有害事象との関連性の検出および評価に影響を与えることについてさらに考慮しなければならない。

データ収集および判定

いくつかの安全性アウトカムでは、イベント特有の症例報告書（CRF）の書式作成が求められる可能性がある。事例（case）の定義には、イベント特有の CRF の情報に加えて、他の情報（例えば、MedDRA 用語、臨床検査値[80]）を含むことが可能である。場合によっては、事例の特定および安全性評価を促進するため、有害事象が疑われる患者に対して、追加の詳細データの収集をプロトコルで求められることがある。例えば、静脈血栓イベント（通常、深部静脈血栓症および肺塞栓症）の特別な定義には、診断の確認フォーム（form）（例えば、静脈造影）が必要となることがある。医師が報告する静脈血栓の日常の有害事象報告は、必ずしも確かな証拠を必要とするとは限らない。特に後ろ向きに情報検索が必要な場合、完璧でない可能性のある医療記録の検索を避けるために、有害事象として静脈血栓イベントの早期発見を目指す際には、確定のための検査結果を収集する特定の CRF を作成することとなる。

ある状況では、判定が必要と考えられる場合がある。例えば、第 II 相試験で指摘された過剰の肝イベントは、第 III 相試験では、判定されたイベントとして表示されることになる可能性がある。そのような場合、専門家もしくは専門家グループ（判定委員会）は、有害事象の医学的分類および標準的な評価方法を提示する。判定は前向きに行われることが理想であり、事前指定と、判定委員会において治療状況の盲検化を継続することは、判定の品質および信頼性をより強固にする。一般に、後ろ向き判定は、前向き判定に比べ困難である。これは、前向きに定義される場合は、判定委員会が求める詳細な医療情報の全てを入手することが容易になるためである。判定が開発プログラム全体で行われる場合は、一貫性を高めるためにプログラム内の全ての試験を通して同じ判定委員会を実施することが賢明である。

3.3.2 後ろ向きメタアナリシスのアウトカム定義

文献のメタアナリシスの大部分は事後に実施される。前向きメタアナリシスで議論された

ものと同じ原則が適用されるが、完了した試験に基づくため、原則の修正が必要になる。運用上の課題を認識し、結果の解釈への潜在的な影響を評価するため、努力すべきである。メタアナリシスのプロトコルにおいて対象アウトカムの明確な定義を指定することは依然として重要である。しかし、試験中に収集されなかった関連情報の欠落により、特定の診断アルゴリズムの実施を評価する能力が、運用上の課題によって妨げられる可能性がある。

理想的には、公表文献のメタアナリシスの基本は、アウトカムに関する周辺状況を理解するために、メタアナリシスのプロトコルには、個々の研究者へ確認する計画を含めなければならない。個別参加者データ (IPD) を利用可能な後ろ向きメタアナリシスの場合、試験で収集したアウトカムの更なる理解およびより良い定義付けのために、IPD を利用することは有用である。

事後の IPD メタアナリシスの第一ステップは、判定、検証と確認努力を計画することである。Hammad らは以下の通り述べている[48]。

主に薬剤の有効性を研究するためにデザインされたランダム化比較試験 (RCTs) は、薬剤の安全性評価のためにデザインされておらず、安全性の検出力を持つことは稀であり、それゆえ、多くの場合、有害事象収集のための体系的で統一された事前に指定された手順が欠落している多くの有害事象の予期せぬ特性が原因で、予め有害事象をしっかりと定義することや、有害事象を特定するために特別にデザインされたツールの使用することは、困難となることが多く、アウトカムの確認および検証のための努力の重要性が強調される。例えば、未成年者[9]での抗うつ剤と自殺傾向[81]との関連性を評価するために実施された抗うつ剤のランダム化比較試験 (RCTs) のメタアナリシスにおいて、研究者は自殺傾向イベントを検索し判定するための評価スキーム[82]を開発した。このスキームにより、メタアナリシスに使用した臨床試験から多くの有害事象が見つかり、再分類されることにつながった。このことは、症例確認方法の相違や臨床経過記述の完全性の相違により、試験間で大幅かつ測定不能な差異が生じる可能性があることを浮き彫りにした。

(自殺傾向 ; suicidality は、現在、自殺念慮および自殺行動 ; suicidal ideation and behavior と呼ばれている。)

スポンサー横断メタアナリシスについては、規制当局、スポンサー、解析を実施する者 (規制当局もしくは第 3 者で) により、プロトコルの全ての内容を協議しなければならない。規制当局あるいは学識経験者から構成された運営委員会であるかは問わないが、統制 ; governance (意思決定機関) の事前合意が必要となる。スポンサーは、一般的に、共通のフォーマットでデータを準備する主要な役割を担い、意思決定の権限は必ずしも持たない

が、解析計画の検討に加わる可能性がある。上述の通り、同じ判定委員会が全ての試験で利用される限りは、一貫性が改善される。個別参加者データ (IPD) が利用可能な完了した試験でのアウトカムの判定の計画は、イベントの分類をサポートできるような文書化レベルを考慮する必要がある。理想的には、明確に定義された臨床基準に基づき診断するために医療記録には十分な情報が必要になるだろう。たとえば、コード化されたイベントが診断の除外を意図していないことを決定するために、最低限、医療コードの正確性をチェックする必要がある。判定委員会の憲章には、判定の決定に使用するための“規則”、特に確証ある臨床データが利用できない場合にどのように決定するか（例えば、医療記録に記載されている治療医師の臨床判断がどの程度信頼できるか）を明確に規定することが必要である。判定の決定は統計解析の結果に重大な影響を及ぼす可能性があることに留意しなければならない。集積されたイベントに真のイベントでないものが含まれる場合、治療による有害反応がマスクされる可能性がある。判定基準が非常に狭い場合、特異性は高くなるが、感度の欠如は、イベントが見逃されることを意味し、バイアスや検出力の欠如につながる可能性がある。

3.4 異なる期間の試験

主要事項の要約

- 一般原則として、治療期間もしくは観察期間の長さはメタアナリシスの除外基準としてはならない。
- 様々な治療期間や観察期間の臨床試験は、リスク増加の時間経過に取り組む上で有益な情報をもたらす可能性がある。
- 要約レベルのメタアナリシスでは、全体的なリスク量に加えて、リスク量と研究期間との関連を評価し、メタアナリシスにおいてその評価結果を提示し、時間の経過に伴うリスクパターンを評価可能なようにすべきである。
- 患者レベルのメタアナリシスでは、全体のハザード比とハザード比の時間依存性を推定するために、生存分析法を使用することが可能である。

一般原則として、試験期間（治療期間/観察時間）は、メタアナリシスの除外基準にしてはならない。特に時間経過にともないリスクが過度に増加すると認識されていない場合、除外基準であってはならない。しかし、先験的な知識や臨床的考察により、特定の医学的な安全性の問題によって、試験期間に制限を加える場合がある。例えば、短い治療期間で副作用を証明すべきではないという生物学的考察に基づく明確な予測がある場合は、メタアナリシスにおいて事前に設定された治療期間より短い治療期間で実施された試験を除外することがある。同様に、急性心筋梗塞後 30 日以内の即時放出型カルシウムチャンネル遮断

薬の死亡リスクに関するクリニカルクエスチョンもある。この場合、急性期の 30 日を経過した後の死亡率は急性リスクをマスクする傾向にある。しかし、薬事規制下の安全性メタアナリシスは、特定の治療に対する有害事象リスク増加の可能性に関する一般的な問題に対処するために実施されることが多い。その観点から、さまざまな治療期間もしくは観察期間の臨床研究は、この問題（クエスチョン）に対処する上で有用である可能性がある。

一旦メタアナリシスに含まれる一連の試験がクリニカルクエスチョンによって決定されると、試験全体を通して推定値を計算する計画とともに、リスクの要約指標（例えば、発生率比、オッズ比、リスク比、リスク差もしくは率差）および 95%信頼区間（CI）の解析計画を作成しなければならない。メタアナリシスが要約レベルのデータに対して実施される場合、リスク指標と試験期間の関連性を評価しなければならない。この関連性は、別々のサブグループでの類似する期間の試験の解析により、もしくはメタ回帰モデルの共変量として試験期間を使用することにより解析可能である。曝露の期間でリスクがどのように変化するかを認知することが解析アプローチに関連することがあることに留意すべきである。例えば、ハザード比が時間によらず一定の場合、長期間の試験の相対リスクは短期間の試験の相対リスクより小さくなるが、長期間の試験のオッズ比は、短期間の試験のオッズ比より大きくなる。このように、試験期間に依存した相対リスクもしくはオッズ比は、ハザード比が期間に依存するというを必ずしも意味しない。

しかし、発生率が小さい場合、薬事規制上の安全性メタアナリシスはほとんどの場合そうなのだが、試験期間により有意に減少する発生率比、相対リスクもしくはオッズ比は期間によるハザード比の減少と一致する。逆のパターンでも、時間経過によるハザード比の増加と一致する。もしハザード比（もしくは増加した他のリスク指標）が試験期間によって異なる場合でも、増加したリスクの要約指標は、イベントと治療の関連性の有効な総合的な検定を生み出す。しかし、全体の要約オッズ比のみの報告は、試験期間がさまざまな場合には不十分である。早期曝露でのリスクの短期間の増加は、長期間のフォローアップ試験と短期間のフォローアップ試験とを混合した要約オッズ比のみの報告によってマスクされてしまう。同様に、長期間の曝露および長期間のフォローアップ後のみ認められるリスクの増加は、短期間の試験と長期間の試験の混合によるマスクされてしまう。試験期間に関連する所見は、解析結果（例えば、試験期間毎の要約オッズ比）とともにメタアナリシス報告書に記載しなければならない。試験期間の関連性は、二次的な所見となるが、リスクパターンを解明するのに有用となろう。

メタアナリシスは、個別参加者データ（IPDs）についても実施され、この場合、イベント時間はイベントを有する患者データから知ることができ、全観察期間はイベントを有さない患者データから知ることができる。これらのデータ（IPDs）は、解析のための観察期

間の選択に柔軟性を持たせる。例えば、全データが利用可能であるため、全ての試験に対して共通の観察期間「T」を設定することが可能である。例えば、T より長い観察期間を持つ患者は、時間 T で打ち切りとする。時間 T 未済までに解析を制限すると、特定の種類のリサーチクエスチョンへの回答に有用であり、脱落の増加など長期間の観察期間に関連する問題を避けることができる。

前述のように、異なる期間のどの試験をメタアナリシスに含めるかを決定することは、クリニカルクエスチョン（長期間もしくは短期間リスク）、作用機序（リスクが治療期間中もしくは治療中止後の延長期間中）、脱落パターン（長期試験の打ち切りによる潜在的バイアス）に依存する。解析のため、共通の期間が選択されようがされまいが、生存率解析の計画では考慮すべき2つの重要な問題が生ずる。一つは、ハザード比が時間によらず一定であることが合理的であるか否か。もう一つは、情報のある打ち切り (**informative censoring**) の可能性である。情報のある打ち切りは、決められた時間で打ち切られた患者が、打ち切られない患者よりイベントのその後のリスクがより高く（もしくはより低く）なる場合に発生する。情報のある打ち切りは、ハザード比の推定にバイアスをもたらす。例えば、癌の研究において、患者は、疾病の進行まで試験薬で治療を受けるが、試験薬を服用する患者が、服用していない患者に比べて、試験薬に関連する有害事象に関して、高いリスクを示す場合に、情報のある打ち切りの可能性は高くなる。従って、標準的な生存率解析は、（少なくとも個々の研究内で）ほとんどの患者がまだ試験薬に曝露されている共通の観察期間を選択しない限り問題がある。臨床開発でしばしば生じるその他の微妙な問題は、薬剤投与中か、そうでないかによらず、観察期間を延長した場合に生じ、治療中のイベントに関連するか、フォローアップ中のイベントに関連するか、公表された文献から決定することが難しいことである。

注意する点がもう一つある。治療期間（もしくはフォローアップ期間）が試験デザインもしくは試験集団の他の側面と交絡となる可能性である。例えば、短期間の治療試験は、重症患者で行われ、長期試験は軽症患者で行われる場合である。このような状況下では、短期試験と長期試験の結果の違いは、治療期間もしくは疾患の重症度が原因である可能性があるため注意が必要である。

3.5 研究の選択とデータの抽出

主要事項の要約

- ▶ 安全性データのメタアナリシスに含まれるランダム化比較試験の選択に影響する主な因子は、対象集団、対照薬、投与量、曝露期間、有害事象の収集方法（能動的または受動的）等の研究デザインの特徴である。
- ▶ 一般的には、これらの特徴が似ている（同一であることは必ずしも必要ではない）研究データを結合することが最も適切である。

3.5.1 研究の選択

メタアナリシスに含める研究の選択は、統合されるエビデンスの基盤を築くことであり、全体的な研究結果に影響を及ぼしうる。研究の選択は、研究結果の性質だけではなく、得られた研究結果の具体的な集団への一般化可能性にも影響する。有効性のアウトカムとは異なり、臨床試験においては有害事象のアウトカムが、事前に十分に構造化され、定義されることはほとんどない。このことが、有害事象のメタアナリシスをより困難なものにしている[83]。

研究の選択には、(3.2節で述べたように)対象集団に基づく考察だけではなく、研究デザインの特徴に基づく考察も含まれる。用量、期間、有害事象の収集方法（能動的か受動的か）及び集団などの特徴を考慮することが重要である。特徴が（同一である必要はないもの）類似している研究データを結合することが通常、最も適切であるためである。治療の（有害）作用が検討した試験の特徴に依存していないときは、メタアナリシスにこのような特徴の類似性は要求されない場合がある。この依存が存在しているかどうかは、部分的には疾病の生物学、薬剤の作用メカニズムに関する知識に基づいた理論的な問題であるが、異質性の項（4.8節）で述べるように、経験に基づいた問題ともなりうる。どの試験のサブセットが結合するのに適しているかを判断する前に、個々の試験プロトコルを徹底的にレビューしておかなければならない。研究プロトコルが似ていることが、研究が似たように実施されたことの保証にはならないことに注意すべきである。プロトコルをどのように実行するかは試験によって異なる可能性がある [48]。例えば、選択基準や除外基準への適合性を確認するために最終的に収集されたデータを調べてみるのが、賢明である[9]。

薬剤との関連性が疑われる有害事象のみの情報を、研究者に記録するよう求めるプロトコルには注意を払うべきである。ICH E3 ガイドライン「治験の総括報告書の構成と内容」は、薬剤との関連性の判断に関わらず、患者が経験した全ての有害事象の報告を研究者に求めている[84, 85]。具体的には、元の研究者が薬剤と関連していると考えた事象に限定するのではなく、報告された全ての有害事象を対象に解析を行うことを推奨する。研究者が

薬剤と関連すると考えた有害事象に限定された研究を含む解析を行わなければならない理由がある場合は、その理由を明確に述べておくべきである。

さらに、製品ライフサイクルのどの時点で研究が実施されたかによって、研究の厳密さも変わりうる。第Ⅲ相で実施された研究と他の相で実施された研究間、あるいは競合品のメーカーが実施した研究間では、測定不能な差異が存在する可能性がある。利用可能な研究の数に応じて、既知の研究目的を含めたデザイン上の特徴で分けて結果を示し、これら特徴の違いがメタアナリシス全体の結果に与える潜在的な影響を吟味することは賢明なやり方であろう。安全性の問題を評価する際に使用されるデータからその問題を示唆したデータを分けて吟味することは、実行可能であれば、効果的な方法である可能性がある。

特定の薬剤研究の中には、異なった用量を含む場合がある。用量反応性を検討するために全用量を含めるか、承認用量に注目するか、過量投与におけるリスクを検討するために承認用量を超える用量（あるいは用量 - 反応を評価するため推奨用量より低用量）を検討するかを決める必要がある。いずれを選ぶにせよ、その選択をプロトコルにおいて文書化し、正当性を示すべきである。

治療の適応症にも、特別な考察が求められる。ある薬剤の全ての適応症をまとめて解析を行うのか、適応症別に解析を行うかを明確にしなければならない。その決定は、治療が有害事象を引き起こすリスクは適応症の間で等しいと信じる理由がある、あるいは、対象の有害事象については、集団（適応症）によって異なる反応を示すと信じる既存の理由があるかに懸かっている[9]。

調査を、ある薬剤クラスの一つの薬剤とするか、クラス全体とするか、あるいは幾つかの薬剤クラスとするかは、早い時期に決める必要がある。決定の合理性は予め明確にしておくべきで、データによって左右されてはならない。決定は、例えば、作用機作や物性、適応症等の類似性に基づく可能性がある。

メタアナリシスが公表文献から利用可能なデータを含む場合、利用可能な全ての研究のシステムティック・レビューを常に先行して行わなければならない。ただし、スポンサーが、市販される前の製品に関するメタアナリシスを行う、あるいは、スポンサーが実施した臨床試験の個人レベルのデータを利用したメタアナリシスを自身が行うといった、特殊な状況では上記は当てはまらない。

「コックラン ハンドブック」[6]等の、有効性データに対するシステムティック・レビューやメタアナリシスの実施に関する公表された専門的なガイダンスには、研究の選択につ

いてのアドバイスが述べられている。幾つかの刊行物は、有害事象や安全性データのシステマティック・レビューやメタアナリシスに対する多様な局面についてある程度議論している[6, 48, 60, 83, 86, 87]。例えば、Cochrane Adverse Effects Methods Group [87]や Safety Planning, Evaluation and Reporting Team (SPERT)では、有害事象や安全性データのメタアナリシスに特化したガイダンスを提供している[60]。他の関連するガイダンスとしては、日米 EU 医薬品既成調和国際会議が発行した ICH M4E [88]、米国 FDA のガイダンス Premarketing Risk Assessment [89]あるいは CIOMS WG VI [59]の報告書がある。

以下のパラグラフでは、Cochrane Handbook[6]において既に十分に開発され、説明されている研究選択のプロセスについて繰り返すのではなく、有害事象の評価における具体的な背景や問題に関し、研究を選択する際、レビュー者が直面する可能性がある諸状況に焦点を当てる。

状況 1：メタアナリシスのための研究セットを作成する方法の一つは、分析結果として得られる治療効果推定量の変動をできるだけ小さくするため、デザイン、投与量、患者集団等が非常に類似した研究だけを集めることである。あるいは、より広く研究を収集し、研究の特徴と結果との関係を分析することにより、デザインや集団の異質性をメタアナリシスの目標として明示的に検討することも可能である（4.8 節参照）。この場合、研究レベルの要因による相互の潜在的な交絡に対し、注意深い配慮が必要となる。例えば、高用量の研究はより重症の患者で実施される可能性が高く、この場合、結果として認められた差が、高用量によるか、疾患の重症度によるかを明らかにすることができなくなってしまう。

状況 2：研究の選択において考慮すべき関連する共通の状況は、同じ化合物を異なった用量で治療している治験を選択する場合である。全ての用量群を結合するメタアナリシスを行っても、通常、あまり有益な情報は得られない。承認された適応症（かつ承認された用量）における有害事象の特性を明らかにするために、プラセボ対照研究の主要な集団から承認用量より高いあるいは低い用量を除いた第一次集団を作成することが理に適っているであろう。次いで、高用量域（すなわち過量投与の可能性）で何が起こるかを評価するために、承認用量より高い用量を加えた第二次集団を作成することがある。一般的に、承認用量より低い用量は、承認用量で起こる効果を薄めてしまうため、除いておくことが推奨される。しかしながら、低用量の研究において起きた事象は無視すべきではないし、用量反応が研究対象となった際には低用量集団も対象となる可能性がある。

異なった用量群を分離するかどうかの判断は、対象とする有害事象が用量に依存する反応であるかどうかによる。例えば、過敏反応やアナフィラキシー反応は非常に低い用量においても起こり得る。このような事象の場合、低用量群を除外することは適切と言えない。

状況 3：安全性データのメタアナリシスにおける問題の一つは、イベントが多く理由により欠落していることである。文献によっては、高頻度に発現したイベントのみを限定して報告するケース（例えば、「我々は治療群において2%以上の頻度で発現した事象のみを報告した」）もあれば、実際にどのような基準でイベントを除外したのか分からないケースもある。イベントを除外する基準はおそらく研究ごとに違っているため、単に報告件数が0件の研究をメタアナリシスに含めるか、除外するかという問題ではない。欠測データに関しては、さらに大きな問題がある。4.2節では、有害事象のない研究の取り扱いの詳細と、各アプローチに関連する限界について述べられている。

状況 4：治療期間もしくは追跡期間も試験間で異なる場合がある。有害事象の種類によって、急激な反応（例えばアナフィラキシーあるいはアナフィラキシー様反応）は、観察期間の早期、おそらく初回投与においてさえ起こり得る。一方、いくつかの事象（例えば、悪性腫瘍）は、おそらく曝露から事象の発現までの間に長い潜伏期間が存在するため、長期間にわたる治療やその後の観察においてのみ認められる。

状況 5：メタアナリシスにおいては、結合される研究の幾つかもしくは全てが、2つより多い群を含んでいるような状況はよくある。多群試験がメタアナリシスに含まれている場合の一対比較 (pair-wise comparison) に対処する方法は、3.7節で説明する。

状況 6：特に治験においては、研究プロジェクトの一部だけが容易に検索できる媒体に公表される。そのため、公表された文献によるメタアナリシスを行うとすると、データは公表されたものに限定されてしまう。研究結果の公表が結果の特性や方向に影響される場合、公表（出版）バイアスが生じる。主に英語と幾つかの他の言語で索引が構成されているデータベースにおいて検索をおこなうと誤った結果が導かれるように、バイアスは公表する際の言語によっても生じうる[90]。3.8節では、公表（出版）バイアスに対処する方法の詳細について論じている。

状況 7：用量変更が可能な試験では、曝露データは最頻の用量で要約される可能性がある。ある用量に忍容性を示した患者がより高用量で治療される状況を考えてみよう。有害事象が起きれば、患者は前の低い用量に戻されるであろう。この患者における最頻の用量は低い方の用量とされるため、有害事象は低用量と誤って結び付けられてしまい、用量効果反応はマスクされてしまう。ちなみに、高用量の投薬を必要とする患者は、通常、疾病の重症度が高く、背景の病状や合併症に関連した有害事象を起こしやすいと考えられる。従って、この状況での用量効果反応の評価では、有害事象が高頻度で発現したのは高用量に起因すると誤って評価する可能性がある[48]。

3.5.2 データの抽出

ランダム化比較試験の成績を公開された文献や他の公開された情報源から入手できても、個別患者レベルのデータを取り出すことはできない場合がある。例えば、個別の安全性懸念事項をより深く理解することに興味がある学術研究者がいるとする（「付録4 ロシグリタゾンの事例研究」参照）。そのような状況では、公表された情報から正確にデータ抽出が行われるように注意を払う必要がある。データ抽出エラーは普通にあることで、多くの研究が報告されており、その幾つかは米国医学研究所のシステマティック・レビューに関する書籍に記述されている[91]。例えば、Jones 等（2005）は、20件のレビュー（48%）におけるデータ抽出エラーについて記しており、その幾つかは（全体の結論に影響は及ぼさなかったものの）結果の要旨を変えてしまっている[92]。

特に正確性と効率性を改善するための方策に関して、データ抽出のプロセスはあまりよく研究されていない。ある研究ではシステマティック・レビューの経験がデータ抽出のエラー率に影響を与えていないことが示された。エラー率は経験に依らず高い値（28.3-31.2%）を示した[93]。データ抽出エラーを少なくするために通常推奨される方策は、2人の人間が独立してデータを抽出することである。パイロット研究において、Buscemi 等は1人のレビューアによるエラー率と2人のレビューアが独立してデータ抽出を行う際のエラー率を比較した。1人によるレビューの場合は、2番目のレビューアが1番目のレビューアが抽出したデータの正確性と完全性を点検し、続いて行われる協議において矛盾を解消した[94]。2人によるレビューでは、それぞれのレビューアが独立してデータの抽出を行った後、同様に協議によって矛盾を解消した。1人のレビューアによるデータ抽出は早かったが、21.7%高いエラー率となった。専門家（例えば、米国医学研究所）は、可能であれば2名によるデータ抽出を推奨している。コクラン共同計画（The Cochrane Collaboration）もこの見解を支持している[6, 第7章]。英国のヨーク（York）大学の The Centre for Reviews and Dissemination (CRD)は、2番目のレビューアによる独立した正確性と完全性についての点検があれば、1人のレビューアによるデータ抽出も容認しうる選択肢であるとしている[95]。

本節では、広く認知され、容易に検出可能な有害作用について良質なデータを提供する、ランダム化比較試験のメタアナリシスを行うことのみならず焦点を合わせていることに注意して欲しい。臨床試験では観察されにくく、新規で、稀で、長期間において発現するような有害作用の体系的な評価には、コホート研究、ケース・コントロール研究、横断研究、さらには症例報告やケースシリーズ等の研究デザインを含める必要がある。3.10節では、観察研究のメタアナリシスに関連した問題をより詳細に検討する。

3.6 研究規模と小規模研究効果

主要事項の要約

- ▶ メタアナリシスの一般原則は、研究規模に関わらず、対象とするリサーチクエスチョンに取り組んだ全ての研究を含めることである。
- ▶ 異なった規模の臨床研究を扱うメタアナリシスにおいては、個々の研究の推定値に与えられる重みづけと、大規模な試験と小規模な試験間での推定値の一貫性が、考慮すべき重要な二つのポイントとなる。
- ▶ 研究の重みづけの選択に基づいた感度分析と、異質性解析の結果に基づく研究の部分集団に対する事後解析を、データの完全な理解を得るために実施すべきである。

研究の選択という広いトピックの下、次に、非常に小規模な研究と大規模な研究を統合する状況に眼を向けよう。大規模な研究（第4章参照）は、もたらす情報量という意味においてメタアナリシスの結果に大きな影響を及ぼすため、この問題は重要である。規制当局の判断は、事実上、一つの臨床研究に左右される可能性があることから、その一つの研究のバイアスがないがことを理解することは重要である。メタアナリシスにおいて優先すべき原則は、対象とするリサーチクエスチョンに取り組んだ全ての研究を含めることである。その理論的根拠は、全ての研究の個別効果量を合計あるいは合成することにより、個別の研究単体と比較して、優れた治療効果の推定が可能となることである。研究が共通の効果を推定しているかどうかに関わらず、この理論的根拠は保持されるが、単一の効果を推定することによって重要な情報を覆い隠してしまう状況もあり得る（4.8節「異質性」を参照）。

この原則の結果として、少なくとも理論的には、研究の選択基準における研究規模の重要性は2次的であると予想される。ある研究が対象とするリサーチクエスチョンに適切に取り組めていないとする理由は多くあると思われるが、研究規模だけではその理由とはならない。実際、メタアナリシスの強みは個別の研究におけるサンプルサイズの問題を克服し得ることである。一方、研究規模は、対象とする事象（イベント）を捉える個々の研究の能力と関係している可能性がある。この章においては、規模というのは情報量のことを指しており、必ずしもサンプルサイズを指してはいない。例えば、対象とする有害事象について高い背景発現率があれば、被験者数が少なめの研究であっても、低い背景発現率で被験者数の多い研究よりも、多くのイベントをもたらす可能性がある。仮に対象とするパラメータが相対リスクであった場合、高い背景発現率を有している状況ではサンプルサイズの小さい研究がより多くの情報を有することになる。イベント数に基づく研究規模の指定

は、治療中の症例数にかかわらず、イベントの総数に基づいて定められるべきである。

異なる規模の研究についてメタアナリシスを行う際に考慮すべき重要な 2 つの点は

- ▶ 多くの小規模試験と 1 つ（幾つかの）の大規模な試験からの情報により総合的な治療推定値を合成する方法。これは、メタアナリシスにおいて個々の研究の推定値にどれだけの重みを与えるかという課題である。
- ▶ 大規模な試験と小規模な試験間での推定値の一貫性。これは結果の異質性の傘下にある特定の問題（対比）である。

メタアナリシスにおいて要素となる研究のサンプルサイズを計画する際には、解析の選択が結果に影響する可能性があるため、計画された解析について検討しなければならない。研究を固定効果（4.3 節 「統計モデルの選択」参照）とみなした場合に用いられる、一般的な重み付けは、対象とする研究固有の推定値の分散の逆数である。（注：4.2 節で指摘するように、逆分散加重は一般的にイベント数が稀な場合はほぼ用いられない。）逆分散加重では全体の推定値が最小の分散をもつようになるが、大規模な試験の推定値により強く影響される。例えば、心血管系に及ぼす効果に関する 14 本の臨床試験のメタアナリシスにおいて、全体のオッズ比の推定に一つの研究が 57%の重みで寄与していた[96]。

先験的 (a priori) に、他の重み付けを選択することも可能である。例えば、大規模試験の影響を小さくするためには、等しい重み付け[equal weights]（典型的には算術平均）が選ばれる。しかしながら、等しい重み付[equal weights]は全体の推定値の分散を大きくしてしまう。小規模試験と大規模試験に等しい重みを付与することが適切かどうかについては議論が続いており、事例ごとに評価すべきである[97]。選択の根拠をメタアナリシスのプロトコルにおいて明確にすべきである。

多くの小規模試験と少数（おそらく 1 つ）の大規模試験からなる場合には、他のアプローチも可能となる。例えば、多くの小規模試験から逆分散法[inverse variance weight]を用いて全体の推定値を得、この推定値と大規模試験の推定値を等加重[equal weight]法により結合させるのである。その結果、小規模試験の推定値と大規模試験の推定値をより直接的に比較することができる。

ランダム効果モデルでは、逆分散加重[inverse variance weights]には、研究特有の推定値に関する研究間の変動の推定値が含まれる。このモデルでは、固定効果モデルよりも、個別の研究に均一の重みを付けることになる。しかしながら、固定効果モデルとランダム効果モデルのどちらにするかは、多くの小規模試験と大規模アウトカム試験の問題に対処するための方法としてではなく、リサーチクエスチョンによって決める（4.3.1 節を参照）こ

とを推奨する。2つのアプローチの間で結果が大きく変わってしまうときは、違いの源が何かを理解するためにより多く努力すべきである。

要約すれば、対象とするリサーチクエスションに取り組んでいる研究は、規模に関わらず、全てをメタアナリシスに含めるべきである。多数の小規模な試験と比較的少数の大規模試験があった場合、全体的な治療の推定値を決定する主要研究の重み付と、異質性の検定をメタアナリシスのプロトコルの方法の項に明記する必要がある。データの十分な理解を得るため、各研究の重み付の選択に基づく感度分析と異質性分析の結果に基づいた部分集団の事後解析を実施すべきである。

試験規模の議論には更に2つのポイントがある。1つは小規模試験と大規模試験の相対的な品質に関するものである。研究規模は品質の高さを示唆している可能性があるが、必ずしも大きなものが良いとは限らない。例えば、Berlin & Kim が抗結核ワクチン「カルメット・ゲラン菌」(BCG ワクチンとして知られている) [98]について述べているように「、、、BCG ワクチンについては、受動的な追跡調査を採用した膨大な数の研究のため、(著者らの認めたところによれば) 結核の症例全体の少なくとも50%を逃し、より完全な追跡を達成できた他の幾つもの研究において確認された予防効果を示すことに失敗している」[99, 100]。Berlin & Kim がさらに、大規模試験は結果の異質性の影響を受けていることを指摘している。彼らは「他の論文[101]では同じ治療に関する大規模試験の間に相違が認められることを指摘している。」と述べている。彼らは、コクランライブラリー[the Cochrane Library]で確認できた同じ治療に関する大規模研究のおよそ27%において、2つの研究結果の間に統計的に有意な差があることを見出している。

これは、メタアナリシスとそれに引き続いた大規模試験を比較したLeloir の広く知られている論文[102]の中で示されている矛盾する比較の割合と同じである。関連した問題が、本質的に探索的性格を有する多くの小規模研究に取り組む、事後メタアナリシスの努力においてみられる。実際のところ、対象とする薬剤と安全性問題の関係について本質を反映しない方向にメタアナリシスの点推定値を圧倒してしまう可能性、すなわち一種の「ボリュームバイアス」が発生する可能性がある[47]。

2つ目のポイントは、規模とは開発プログラムの規模を指す薬事規制の観点に由来するものである。対象とするリサーチクエスションが、ある薬剤クラスの安全性に関連する場合がある。クラス効果の推定値を得るために開発プログラムの中での研究に対する重み付けや、開発プログラムを跨いでの重みの割り当てには、開発プログラムの規模を考慮に入れるべきである。より詳細な議論は幾つかの事例(例えば、選択的セロトニン再取り込み阻害剤(SSRIs)と自殺傾向[9]、抗てんかん剤と自殺傾向[103])において確認することがで

きる。クラス内の薬剤間での潜在的な結果の異質性によって、特に開発プログラムが小規模で、対象とする特定の有害事象を検出する可能性が低い薬剤が存在する場合、クラス効果を推定することが適切であるかが問題となる場合がある [48]。このように、薬剤クラスに対するメタアナリシスにおいては、1つの薬剤を対象とするよりも規模の問題は複雑になる。

3.7 多群試験

主要事項の要約

- ▶ 多群臨床試験には、二つ以上の介入群と一つの共通の対照群がある場合、あるいは、プラセボ群と標準治療群のように複数の対照群がある場合がある。
- ▶ 多群が正しく扱われない場合、対照群のダブルカウントが起きてしまい、ある結果が過剰表現され、メタアナリシスにバイアスを生じさせてしまう。
- ▶ メタアナリシスに多群試験が含まれる場合、一対比較を扱う対処方法は、事例ごとに、その状況に応じて決めるべきである。
- ▶ 単純な一対比較とするために、群をまとめることが推奨される場合もある。一方で、高用量と低用量をまとめることによって高用量の効果を希釈し、低用量の効果を誇張することがある。

多群臨床試験には、二つ以上の介入群と一つの共通の対照群が存在する場合や、プラセボ群と標準治療群のように複数の対照群が存在する場合がある。多群比較試験をメタアナリシスに含める場合に直面する共通の問題の一つが、一対比較を扱う対処方法である。

多群研究は、二つの介入群（介入 1 と介入 2）と一つの対照群からなる 3 群並行試験であることが多い。二つの介入群は試験的治療の低用量群と高用量群などとなる。それぞれの介入群を別々に解析に含める場合（例えば、「低用量とプラセボの比較」と「高用量とプラセボの比較」の両方を同じメタアナリシスに含める場合）、対照群は二度解析に使用されることとなる。このような場合、過剰な重み付けという「解析単位」の問題が生じる。この問題は対照群のダブルカウントと言われるもので、特定の結果が大きい比率を占めることとなりメタアナリシスにバイアスを生じさせる[104]。

二群より多い群が関連する多群比較試験をメタアナリシスに含めるための方法に、唯一の正解はない。事例ごとに、それぞれの状況に応じて決めるべきである。多くの状況に対応できる可能性のある方法の一つは、群をまとめて一組の一対比較とすることである[6]。し

かしながら、高用量と低用量をまとめることによって、高用量の効果を希釈し、低用量の効果を誇張させることがある。両用量群を別々にメタアナリシスに含めることができる方法（次のパラグラフを参照）であれば、用量反応相関を評価することが可能となる。

様々な介入群を組み合わせる方法には、評価されるアウトカムの種類によって異なる幾つかの方法がある。研究の主たる安全性情報が、各群の事象を発現した患者数（あるいは事象の発現件数）と患者総数（あるいは患者年）である場合、多くの出版物において事象の発現件数そのものが公表されている。そのような状況では、二つの群を組み合わせて、一の群として取り扱うことで、ダブルカウントを防ぐことができる。二つの群の事象発現件数と患者総数（あるいは患者年）をそれぞれ足し合わせるによって行われる [105]。時には、対象となる健康上のアウトカムは連続変数で、評価される主たる安全性情報は各群の患者総数、アウトカムの平均値や標準偏差となる。

このような状況では、複数の介入群の平均値や標準偏差を組み合わせて、試験を複数の治療群にまとめて再構成することができる [6]。しかしながら、このアプローチはデータに関して誤った結論を導く可能性がある。例えば、プラセボ群と標準治療群の 2 つの対照群、あるいは、異なった薬剤を用いる実薬対照群を複数もつ多群試験があるとする。プラセボと様々な標準治療が対象とするアウトカムに異なった作用を有していた場合、それらの群を結合させて一組の対比較を実施することはバイアスを生じさせるだろう。これは、先のパラグラフで議論した、同じ試験的治療の用量が異なる群をまとめる際に生じた問題と、同様の問題である。2 つの異なった実薬対象（低用量と高用量のように）が事象のリスクに対して異なる効果を有している場合、それらを結合した群から導かれる結論を解釈することは困難である。

ダブルカウントによる「解析単位」の誤差を克服する更なる可能性は、その解析における同じ研究の相関を比較したもの間の相関を説明することである。メタアナリシスに使用される階層ベイズモデルは、同一研究内での補正された比較を組み込むことができる。Gieser & Olkin は頻度論的な代替案を提示している [106]。

多群比較試験の議論では、もうひとつの考慮すべき別の問題、すなわち、全ての研究の全ての群をメタアナリシスの対象として使用するか否かの判断の問題を生じる。簡単な例として、特定の非ステロイド系抗炎症薬（NSAID）に関連した、プラセボと比較した際の心血管イベントのリスク増加をリサーチクエスチョンとしていた場合、ある NSAID をプラセボに加えて別の NSAID と比較していた研究では、二つ目の NSAID はメタアナリシスから除外することが適切な選択となる可能性がある。もし、クラス効果を検討しているのであれば、その決定は適切とはいえない。介入薬剤と各比較薬とを別々に評価する必要がある

あるならば、混合治療比較を用いない限りは、それぞれの比較は別のメタアナリシスで行われるべきである。

3.8 メタアナリシスのバイアスリスク、個々の研究と研究全体

主要事項の要約

- ▶ ランダム化比較試験が薬剤の安全性評価のために、特別にデザインされ、検出力が設定されることは稀である。有効性の主要エンドポイントの評価において研究の妥当性に影響を及ぼさないデザイン側面であっても、安全性評価の妥当性に影響する可能性がある。
- ▶ ランダム化比較試験は多くの場合、早期、高率に被験者が試験から脱落してしまうことにより毀損され、特に観察期間（follow-up）が長期にわたるものにおいては脱落バイアス(attrition bias)が導入される可能性がある。
- ▶ 特に、有害事象の確認や判定の計画が事前に定められていない場合においては、薬剤との関連が知られている有害事象の発現が振り分けられた処置を明らかにしてしまい、バイアスを導入する可能性がある。
- ▶ ランダム化比較試験における被験者集団の充実は、通常は、患者の参加と応答を改善し、治療の長期のベネフィット評価を可能とすることによる試験効率の改善を意図しているが、実際には、薬剤と関連する有害事象の発現率を過小評価する場合がある。
- ▶ 稀な有害事象を調査する際、薬剤クラス内の異なる薬剤のランダム化比較試験を結合することが必要となることがあるが、要約推定値の解釈可能性は、有害事象の発現率の増加が薬剤クラス内では同等であるという仮定に依拠している。

メタアナリシスから得られる情報の質は、含まれている研究の妥当性に強く依存していることは明らかである。このことは疑いようが無いように思えるが、安全性評価という状況においては微妙な意味合いを持つ。ランダム化比較試験が薬剤の安全性評価のためにデザインされ、検出力が設定されることは稀である。主要エンドポイントの有効性評価では、研究の妥当性に影響を及ぼさないデザイン側面であっても、安全性評価の妥当性に影響する可能性がある。安全性評価の際に遭遇することが多い稀なアウトカムの研究においては、潜在的な研究の限界が悪化する可能性がある。状況によっては、ランダム化が完全には成功しない場合があり、これにより、重要なリスク因子の永続的な不均衡が起り、未調整の推定値のみ用いられる場合には、相対リスク推定値の妥当性に影響する可能性がある。理想的には、メタアナリシスは前向きに計画されるべきである。しかしながら、幾つかの

メタアナリシスで見受けられる事後的な性質により、後ろ向き観察研究に固有の潜在的なバイアスが導入される[27, 107-110]。

ランダム化比較試験の「品質」を評価する複数のツールが提案されている[111-113]。Olivoらはランダム化比較試験の方法論的な品質評価に使用される尺度[scale]をレビューし、尺度のほとんどが適切に開発されていないと結論付け、メタアナリシスにおいて品質の尺度による試験の重み付けをしないよう勧告した[114]。Juniら[115]は研究の品質と研究の結果との関係は、品質の指標 (measure) に大きく依存していることを示した[116]。臨床試験のデザインや運営の個々の側面の方が、各項目から得られる恣意的に重み付けされた要約スコアよりも、バイアスの潜在的な原因（供給源）を調査するのに有用であるとの見方が一般的となっている[116, 117]。しかしながら、バイアスの原因（供給源）を評価する際の挑戦課題は、公表されたメタアナリシスの情報不足、すなわち、メタアナリシスを構成する各研究の薬剤安全性に関連する可能性のある種々の側面についての情報不足である[48]。本節では、メタアナリシスに含まれるランダム化試験の「妥当性への脅威」と「バイアスを生む機会」とともに、これらの試験を結合する試みについて述べる。

3.8.1 早期試験中止によるランダム化の脆弱性

実際には、ランダム化比較試験は、早期に、高率に被験者が試験から脱落することにより毀損されることが多く[118]、これにより、脱落バイアス[attrition bias]を導入する可能性がある。被験者が脱落する可能性は時間の経過と共に増大していくため、長期の経過観察試験において特に問題となる。例えば、抗うつ薬と青少年の自殺との関係を検討するメタアナリシスの場合、試験の中止率は8%から66%の幅が有り、幾つかの試験においては比較する群間で中止率に有意な差が認められた。慢性閉塞性肺疾患（COPD）患者における抗コリン吸入剤と主要心血管イベントのリスクを検討するメタアナリシスにおいても、幾つかの臨床試験で相当な試験中止率（42%に至る）がみられた[15]。既知の測定された特性については不均衡を調整することは可能だが、測定されなかった特性の不均衡を調整することは困難である。観察されなかった交絡因子の潜在的な影響に対して、観察された関連性（associations）の感度を評価するため、様々な方法が提案されている[119-122]。中止理由が対象とする有害事象の発現率にも関連する場合、更なるバイアスが導入されることになる。例えば、胸痛があり、心筋梗塞につながるリスクが高い患者が早期に薬剤群から脱落していくと（臨床的見地からはおそらく適切であるのだが）、薬剤の曝露と心筋梗塞との関係は検出できない可能性がある（すなわち、薬剤群の患者脱落によって、人為的に心筋梗塞のリスクが減少する）。別の例では、抗コリン薬の吸入剤がCOPDの悪化を抑制していることを示す6ヵ月間のランダム化比較試験における死亡率をレビューしたレトロスペクティブ研究がある。この研究では、試験治療の早期中止がプラセボ群で15%、薬剤

群で 8%起きている。重要なのは、より劣った治療（すなわち、プラセボ投与）を受けた重症患者（すなわち、より肺機能が低下した患者）は試験中止の可能性が高まることである（すなわち、脱落によりプラセボ群の死亡率は人為的に低下し、薬剤群がより悪くみえることになる）[123]。

試験からの脱落理由や、試験を中止した患者について健康や生存の状態に関する情報を通常は満足できる詳細さで捉えることができないため、研究群間に生じた不均衡を制御することは困難である。このように、バイアスがどちらの治療群にも好ましく働くことがあることを知っておくことは重要である。例えば、仮の話として、抗コリン薬の試験において、積極治療により症状が緩和した重症患者は、プラセボ群の重症患者よりも長期間の試験を継続し、薬剤曝露とは無関係に、根底にある疾患の重症度が増すことによって、有害事象の発現リスクが高くなる可能性がある。

3.8.2 割り当ての非盲検化

マスキング（盲検化）の失敗は有害事象の評価の妥当性をも損ないかねない。特に、有害事象の確認や判定の計画が事前に定められていないような場合には、薬剤に関連することが知られている有害事象の発現が、割り当てられている処置を明らかにすることとなり、バイアスを導入する可能性がある。あるクラス全体の薬剤に有害事象が関連すると知られている場合（例えば、抗うつ薬における性機能不全）には、特に困難である[124]。

非盲検化は有害事象の認知と観察の増加につながり、最終的にプラセボより薬剤が投与されている患者で他の有害事象が検出される可能性が高まることになる。同様に、薬剤に関連した有害事象はより頻繁に医療提供者との接触する場合がある。例えば、女性の健康イニシアチブ研究においては、ホルモン補充療法（**hormone replacement therapy: HRT**）に関連する膣出血の発現とその管理がプラセボ群より薬剤治療群に多発したことによって、盲検性は早期に破られてしまった[125]。他の例としては、尿が着色してしまう薬剤における、監視（検出）バイアスに起因する尿道がんの見かけのリスク増加が含まれる可能性がある。同様に、乳房の圧痛をもたらす薬剤においては、乳がんを診断する検査の頻度が増し、その後の明らかな（しかし、誤った）乳がんリスクの増加をもたらした可能性がある。

参加者の登録や割り当てに関わる人間に対して、割り当ての順番が適切にマスキングされず無作為化の順番が保護されていない場合、無作為割付けの前に、割り付けが明らかとなる事態が起り得る。次の割り当てを知っていることにより、リスク因子を基に参加者をえり好みして登録してしまう可能性が高まる。割り当てマスキングの失敗は薬剤の有益性を過大評価する傾向があり、薬剤の安全性に関連する重大なバイアスを導入する可能性がある

ある。

3.8.3 有害事象の確認

有害事象の確認が一貫していないことに関連した様々な問題は安全性の評価を毀損しかねない。例えば、有害事象の先験的 (a priori) な定義や、特別にデザインされた有害事象を確認するための方法 (tool) を前もって定めていないことは特に問題となる。どのような有害事象が重要となるかを常に予想しておくことはできず、起こり得るすべての有害事象の定義を予め定めておくことも現実的ではないため、引き続き実施する試験において安全性の収集法方法を変更できるように、(第Ⅲ相の終わりに限らず) 開発の過程において安全性データを定期的に評価することが重要となる。しかし、有害事象の収集方法を変更しようとする際には、将来、実施されるメタアナリシスへの影響も考慮すべきである。例えば、有害事象の収集方法が変更される前後の試験について、解析が分けられて計画され、実行される場合がある。この分けた解析からそれぞれ結果が得られる。先験的 (a priori) に変更後のデータがより確からしいとみなされ主解析とされたならば、有害事象の定義の特異性が向上したことで、リスクの増加分がより大きくなる限り、リスク増加に対する統計的検出力は限定される可能性がある。

臨床的記述の完全性が試験間で異なっていると、試験間の比較は困難となる。さらに、薬事規制を目的とせず作成されたランダム化比較試験の報告書や、公表文献では、全ての有害事象が記載されていない (出版バイアスの一形態)。公表された試験の報告書には有害事象については記載がないことが多く [126]、可能であれば、元の調査報告書や公表文献には含まれていない情報を入手するために元の研究者に連絡することの重要性が強調されている [127]。抗コリン剤と心血管イベントとの関係性を評価するメタアナリシスにおいては、研究対象とする心血管系の有害事象について記載がない試験は、おそらくイベント発現率が低い可能性があるとの推測から対象外とした結果、利用可能な試験のおよそ 3 分の 2 が除外された [15]。もし、除外した試験においてイベント率がバランスしていなかったとしたら、多くの試験を対象外としたことでメタアナリシスの結果に何らかのバイアスが導入される可能性が高い。メタアナリシスにおいて示唆された心血管イベントのリスクの増加は、その後の大規模で、長期間のプラセボとのランダム化比較試験において否定された [128]。(注意：有害事象について適切に評価・報告された研究と、どちらの治療群にもイベントが発現していないことが明確に報告されている研究の取り扱いが、関連する問題として 4.2 節で述べられている。)

ランダム化比較試験の被験者集団を濃縮 (enrichment) することは、通常、患者の参加と反応を改善することによる試験効率の改善を意図するが、ランダム化比較試験からの医薬

品安全性データの解釈に影響を及ぼす可能性もある。例えば、無作為化脱落デザインにおいては、始めの導入期に薬剤に反応しなかった、あるいは忍容性を示さなかった患者は通常試験から除外される。この戦略は試験の効率を高め、長期の治療有効性の評価を可能とするが、薬剤に関連した有害事象の発現率を過少評価する可能性がある[129]。抗うつ薬の試験において、始めに登録された 533 名の患者のうち、薬剤に反応した（そして、おそらく忍容性を示した）280 名が、引き続き二重盲検の治療に無作為に割り当てられた。有害事象による中止率は二重盲検の治療期では薬剤群とプラセボ群ではほぼ同じ値であったが、導入期において薬剤に曝露された患者のうち 61 名（24%）は有害事象の発現のために試験への参加の中止を余儀なくされた。無作為化前の期間の有害事象発現率がどれほど重要な情報を含んでいたとしても、薬剤の安全性評価においては、検討されないことが多い[130]。同様に、有害事象の過小評価はクロスオーバーデザインにおいて起こる、患者が比較する治療群に切り替える際に、第一期の薬剤に忍容性を示さなかった患者は脱落するため、第二期に参加することはできなくなるからである。

3.8.4 個別試験のデータ収集

メタアナリシスに関して頻繁に懸念されることは、研究結果の性質による選択出版、すなわち、出版バイアス（Publication bias）の可能性である。試験全体が論文化されていない場合は、利用可能なエビデンス全体の評価を試みるメタアナリシスの妥当性に対して、重大な脅威となる可能性がある。[131]。

メタアナリシスにおいて、ファンネルプロットは試験結果の分布の非対称性の確認に使用され、非対称の分布は、出版バイアス（Publication bias）の可能性を示唆する[132]。ファンネルプロット（Funnel plots）は、試験毎のサイズや制度の測定に対する個別の試験による介入効果の推定量の単純な散布図である。バイアスが存在しない状況では、図 3.1 に示されるように、プロットはおおよそ対称な漏斗（Funnel）に似ているはずである。このグラフのポイントは、試験サイズの指標（縦軸上に標準誤差のログオッズ比として提示）に対する治療効果（横軸上にオッズ比として提示）をそれぞれ個別の試験毎にプロットすることである。小規模試験では大きい標準誤差となる傾向があり、これらはグラフの下端に表示される。しかしながら、例えば、統計的に有意な効果を示さなかった小規模試験（図 3.1 で白抜きで示されている）が公表されないバイアスが存在する場合、グラフの下端にギャップのある非対称な外観のファンネルプロット（Funnel plot）へつながる。

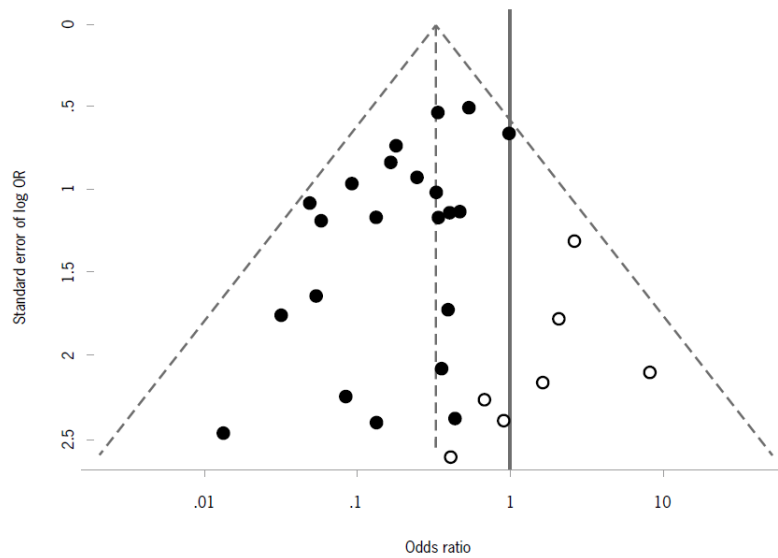


図 3.1 仮想的 Funnel plot: バイアス非存在化における対象なプロット

出典: Figure 10.4.a Cochrane Handbook for Systematic Reviews of Interventions [6].

以下の図 3.2 は、心筋梗塞による致死率に対するマグネシウムの効果を検討した 15 本の試験結果の事例である[133]。このメタアナリシスでは、小規模な試験の多くは死亡リスクの低減を示したが、続く大規模な試験では確認されなかった。その結果、非対称の大きいファンネルプロットとなった。非対称の Funnel plot を自動的に出版バイアスと考えるべきではないことは知っておくべきだろ。なぜなら、非対称性は別の要因、例えば手法の品質の違いによっても引き起こされるからである。ファンネルプロットを解釈するに際しては、ファンネルプロットの非対象性に関する考えられる様々な理由を区別することが不可欠である。

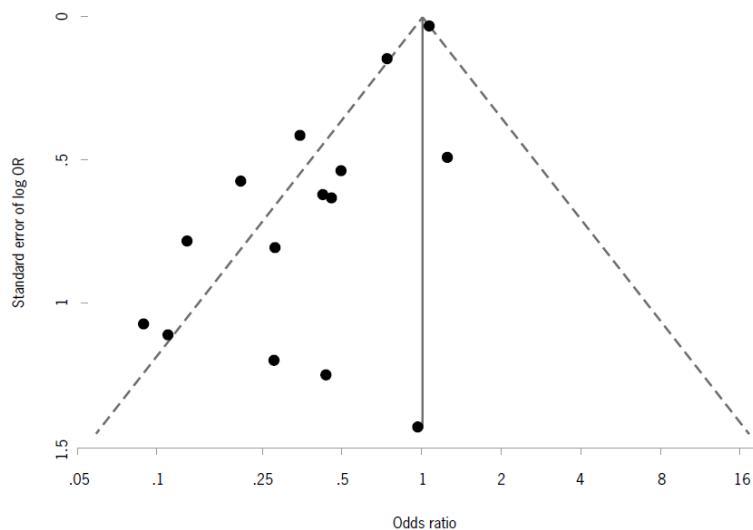


図 3.2 バイアス存在下における非対称なプロット

出典: Sterne & Harbord. Funnel plots in meta-analysis. Stata Journal, 2004, Figure 2, p.131 [133].

近年の Clinical trial.gov への比較試験の登録要求は、より多くのデータがメタアナリシスで利用可能となるため、研究レベルで報告を選択することの影響を低減する可能性がある[134]。同様に、欧州と米国の企業団体がアナウンスした方針

(<http://phrma-docs.phrma.org/sites/default/files/pdf/PhRMAPrinciplesForResponsibleClinicalTrialDataSharing.pdf>)では、薬事規制下の研究報告の全て、更には個別の参加者データまで、それを要求する科学者に対して提供する義務がスポンサーに付されている。これには、出版バイアスの低減が意図されており、他では報告されなかったエンドポイントの解析までも可能とする（そして、最初からより完全な報告書の提出が促進される可能性がある。）

出版バイアスの潜在的な影響の例として、cefepime の使用に関連した死亡率上昇の懸念を提起するメタアナリシスが公表された。しかしながら、全ての試験を包含していなかったことが研究結果を変えてしまった[135]。このシグナルを調査するために、米国の FDA は 1 年間に渡る取り組みを行い、全ての cefepime の公表された試験結果と公表されていない試験結果を調査した[10]。FDA の解析では元のメタアナリシスが 57 本の試験なのに対し、88 本の試験を含んでいた。元のメタアナリシスで報告された致死率の上昇は、FDA が行った更に大きくより詳細なメタアナリシスでは確認されなかった。

対照的に、主要な医学雑誌を含む文献に複数の事例を見つけることができるが[15, 136]、試験や参加者の多重カウントの問題については、あまり注目されなかった[104]。そのため、異なる出版物が異なる研究集団を実際に反映することを確実にするために、臨床試験の実施者とコミュニケーションをとることが重要である。

3.8.5 併合のための試験の適合性

研究が医学的に類似している場合、すなわち「同じ」クリニカルクエスチョンに焦点を当てている場合、試験間のデータ併合は適切であり、治療効果はランダム化比較試験間で相対的に類似することが期待される。しかしながら、「類似する」あるいは「同じクリニカルクエスチョン」のような言葉を定義する事は難しい。試験のデザイン、プロトコルのニュアンス（微細な違い）または実装、薬剤の適応症、選択基準や除外基準、治療の期間や投与量、そして有害事象について症例確認方法の違いが研究間における結論の差異に繋がり、要約推定量の解釈を困難にする可能性がある。3.5.1 項で述べたとおり、試験が類似の初期プロトコルを持つという事実だけでは、試験の類似性は保証されない。プロトコルの実際の実行がそれぞれの試験で異なる可能性がある[48]。一つの薬剤クラス内で、異なる薬剤に関するランダム化比較試験を結合する事は、稀な有害事象を評価するために必要となる

可能性があるが [9]、要約推定量の解釈が可能かどうかは、そのクラス内において、特定の有害事象の発現率の上昇が類似しているという仮定に依拠している。この仮定は、妥当ではないかも知れず、特に稀な事象の場合は、これを証明したり否定したりすることは困難なことが多い。更に、何年もかけて実施された試験を併合するメタアナリシスでは [9]、改善された治療や新たな技術が利用可能になった事による医療行為の変化が試験間の変動性を増加させる可能性がある。

3.8.6 分析単位の問題

コクラン・ハンドブックには、以下に再掲する分析単位の問題の有用な概要が掲載されており、分析単位の問題が一般的に発生する状況のより詳細な一覧と関連する議論の手引きが記載されている [6]。

メタアナリシスに組み込まれる臨床試験に関する重要な原則は、分析レベルとランダム化レベルが一致する必要があることである。実際には、これはしばしば（いつもというわけではないが）分析での観察数はランダム化された「単位」の数と一致すべきということを意味している。より一般的には、いわゆる「デザイン効果”design effects”」の調整を分析に適用する必要がある。結果として、ランダム化された単位数と単位内の個体の総数の間にある一種の有効なサンプルサイズをもたらす。

例えば、単純な並行群間デザインにおいては、各個人が 2 つの介入グループの 1 つにランダムに割り付けられ、各参加者から各アウトカムについて単一の測定値が収集され分析される。しかしながら、実際のランダム割り付けのアプローチは様々であり、ランダム割り付けの方法は考慮すべき重要事項である。研究の著者は、各研究で以下の点について考慮すべきである。

1. 個体群をまとめて同じ介入へランダムに割り付ける（クラスター-ランダム化試験 cluster-randomized trials）
2. 複数の介入を各個人にランダムに適用する（例、クロスオーバー試験あるいは各個人を複数施設で同時に治療）
3. 各個人に 1 つの介入を適用しても、同じアウトカムに対して複数の測定値がある（例、繰り返し測定、事象の再発、異なる身体部位の測定）

3.9 個別参加者データ、要約レベルデータまたは両者へのアクセス

主要事項の要約

- ▶ 従来、メタアナリシスの多くは、伝統的に要約レベルデータ（通常は公表データを意味する）を基にしてきたが、個別参加者データへの関心が高まっている。
- ▶ メタアナリシスの背景情報（コンテキスト）に個別参加者データを使用する利点として、次の事項が可能となることがあげられる。
 - － 全てのデータを共通の定義セットで調査し、試験間の用語の一貫性を向上させる（例：貧血として定義されるヘモグロビン値の標準化）
 - － アウトカム定義における複合変数（**composite variables**）の使用。
特定の事象を定義する変数を共通の生理学的機序を示す変数の組み合わせに基づいた複合変数とする。例えば、有害事象（例：食欲抑制）を体重減少または食欲減衰の組み合わせとして定義することが可能となる。
 - － 時間事象(事象までの時間 **time-to-event**)解析の実施。
単純な割合として表示する要約レベルデータでは実施困難である。
 - － (要約レベルとは対照的に) 参加者レベルの特性によって定義される参加者のサブグループにおける治療効果の推定。
 - － 対象群間における不均衡の調整。
(例：個別参加者レベルの様々な人口統計学的データや合併症を考慮する)
- ▶ 要約レベルデータと個別参加者データの組み合わせが利用可能な状況に合わせた手法が開発されている。
- ▶ 一部の者は、メタアナリシスは観察デザインを象徴すると考えている。そのため、メタアナリシスのアウトカム、推論、解釈は因果関係として報告するのではなく、関連（**associations**）として記述すべきであると考えている。

多くの公表されたメタアナリシスは要約データのみを基にしているが、個別参加者データの使用には利点がありそうである。実際には、重要な分析の原則（例：試験毎の層別化）は患者レベル及び要約レベルの分析の両方のアプローチに適用される。

要約レベルデータのメタアナリシスで可能な分析は、個別参加者データのメタアナリシスでも実施可能であるが、その逆は真ではない。個別参加者データの利点を以下（3.9.1 項）に示す。これらの利点から、薬剤開発において、依頼者が関連する試験とそれに関係する個別参加者データの全てに対してアクセス可能であるときに、なぜ個別参加者データが常に利用されるわけではないのかを問うのは理にかなっている。これには、少なくとも 2 つの重要な理由がある。第一の理由は、分析データベースを作成するために必要な統合作業

は、特にレトロスペクティブに実施する際は、多大な労力が必要となることである。また医薬品のクラスエフェクトが安全性の課題である際は、異なる有害事象の定義や異なる追跡期間を使用している多くのスポンサーからデータが提供されることから、統合はより困難で時間がかかる。第二の理由は、時に個別参加者データは、その薬剤を市販しているスポンサーであっても、幾つかの試験を利用できない可能性があることである。例えば、既存医薬品の新効能のための試験は、患者レベルのデータを共有しないアカデミア (academic cooperative group) が実施した可能性があり、あるいは、他のスポンサーによって当該薬剤がアクティブコントロールとして利用された可能性もある。重要なことは、個別参加者データの使用に必要な追加の労力について、どのような場合に価値があるかを考慮することである。例えば、個別レベルデータのいかなる面においても再評価が必要ない場合（例：イベントまでの時間の計算、薬剤群と比較群の中止の差の検証あるいは有害事象の記録）や、取り込まれる課題が患者特性によって定義されたサブグループを含まない場合は、要約レベルデータのメタアナリシスに対応する個別参加者データのメタアナリシスと同じ回答を与える[137]。

重要なことは、要約レベルデータのメタアナリシスに適用される多くの制限事項は、個別参加者データのメタアナリシスにも適用されることである。PRISMA の IPD（個別患者データ）ガイダンスに付随する論説では「データ共有への関心の高まりは、IPD メタアナリシスの普及を招くため、この試験デザインは、ますます著者と読者に関係していくであろう」と言及している[19]。興味深い事に、米国医学会雑誌（JAMA）で発表された JAMA 編集者の一人が共著者である論評において、「PRISMA-IPDの声明で提案されているように、JAMA はメタアナリシスを観察的なデザインに相当すると考えており、メタアナリシスの結果や推論、解釈は、効果の大きさ（size of effect）のような因果関係として報告するよりは、関連（associations）として記述すべきである」と特に言及している[25]。

3.9.1 個別参加者レベルデータの利点

データ基準の変更は、全ての医薬品の開発期間中においてあり得ることである。単純な事例としては、コーディング・システムや辞書の変更や更新が考えられる。有害事象が新しい版の辞書（例えば MedDRA）[80]に基づいて報告される際は、同じ用語が異なる基本語に割り当てられる可能性があり、また上位階層で差異が生じる可能性もある。個別参加者データへのアクセスにより、全てのデータを共通の版の MedDRA へ割り当てることができるので、試験間の用語と定義の一貫性を向上させることが可能となる。一般に、個別参加者データへアクセスすることで、全ての試験で共通の変数を創出することが可能となる。例えば、結果の層別化のための年齢区分は、異なる基準や境界を用いて、試験ごとに既に決められている。また、貧血を定義するために使用されるヘモグロビンの閾値が試験ごと

に異なる可能性がある。このような状況では、患者レベルデータへのアクセスにより、要約レベルデータのみでは不可能な試験間の標準化が可能となる。個別参加者データを使用することで、中途打ち切り事象 (censoring events) の定義も標準化できる。これもまた、要約レベルのデータでは不可能である。

患者レベルのデータが利用できると、アウトカムを定義するために複合変数を使用することが可能となる。複合変数は、特定の事象を定義する各変数を複数組み合わせ、集合的に使用することによって、共通の生理学的機序を表すものとして用いられる。例えば、有害事象を、体重減少あるいは食欲減衰の組み合わせとして定義することが可能となる。また、同じ根本的な問題の異なる指標の一致 (concordance) をチェックすることが可能となる (例: 有害事象としての体重減少あるいは体重の指定された閾値を越えた減少)。さらに自殺評価のコロンビア分類アルゴリズム「the Columbia Classification Algorithm of Suicide Assessment (C-CASA 基準)」[82]に基づく自殺事象のランク付けなどの場合などで、判定 (adjudication) が必要となるアウトカムの事後解析を実行できる。事後判定 (post hoc adjudication) は、研究全体を通して定義を標準化する際や、研究を通してどのように操作されているか、どのように各試験の治療効果を比較可能 (必ずしもより類似させる必要はない) とするかなどの点で、重要な役割を担っている。

患者レベルのデータは、事象までの時間分析 (time-to-event analysis) を可能とする。単純な割合の形で表示されている要約レベルのデータでは事象までの時間分析は、簡単には達成できない。患者レベルのデータと、Kaplan-Meier 「事象までの時間 (time-to-event)」グラフは、時間の経過と共にハザード率が変化する事象 (例: 長期の追跡で認められなくなる初期の曝露に関連した急性の事象) や潜在的誘導期間を有する事象 (例: 悪性腫瘍) を見出すのに役立つ可能性がある。個別参加者データがあると、生存解析のためのいくつかの重要な前提を確認できる可能性がある。ハザード率の時間依存性 (例: 時間が経過してもハザード率が一定かどうかの確認) は個別参加者データを使用することで調査できるし、調査すべきである。これは、Cox model における治療時間相互作用による時間、T を超える離散単位時間内の log-rank 分析、および/または Kaplan-Meier 曲線の検査により推定可能である (ただし、複数の試験が併合された Kaplan-Meier 曲線は誤解を招く可能性があり、試験毎に確認することが重要である)。これらの試験固有の相互作用と分散は、研究全体の相互作用の要約指標の算定において利用可能である。ある事象のリスクを高めることが知られている薬剤の場合、相互作用自体がメタアナリシスを実施する主な理由となる可能性がある。すなわち、曝露開始からどの程度の時間でリスクが高まるのか、あるいは低下するのかを理解することが関心事となる。

個別参加者データの重要な利点は、被験者レベルの特性により定義された患者のサブグループ

ープにおける効果を推定できる点である。例えば、Szczeczek らの分析では、腎移植の失敗リスクの高い患者のサブグループにおいて、導入療法（基本的に、移植腎臓への拒否反応を防ぐ予防治療）の特定効果を検討することが出来た [138]。これは、厳密には安全性の問題ではないが、この事例は、患者レベルデータへのアクセスにより得られる柔軟性について潜在的な重要性を示している。Szczeczek らは、要約レベルデータだけでは特定できなかった治療と患者レベルの特性との関係を検出することができた。具体的には、対象とする共変数は、免疫システムの感作指標であるパネル反応性抗体レベル (the panel reactive antibody level (PRA)) であった。

2年間のフォローアップの後、著者 (Szczeczek) らは個別参加者データを用いる事で、導入治療の効果が感作された患者とそうでない患者で異なることを示した (相互作用に対して $P=0.03$)。この相互作用は、5年のフォローアップ時点でもまだ有意であった。逆に、試験レベルの特性として「感作された患者の割合」を使用し、治療効果を試験レベルの共変数に対して回帰推定した要約レベルデータの分析では、この相互作用について、一貫しては、証明されなかった [139]。独立したデータセットにおいて、そのサブグループの結果を確認することは重要である。なぜなら、もしサブグループでの効果が確認されれば、高い効果が期待できる集団に対して導入治療が実施される一方で、他の患者では不要な治療や潜在的な毒性の可能性が回避されることを意味するからである。

サブグループでの効果を検出するために要約レベルデータを使用することには、他にも限界がある。例えば、Lambert らは、要約データに基づく分析では、相互作用を検出するための検出力 (Statistical power) が低いことが多いことを示した (effect modification としても知られている) [140]。同様に、Schmid らは、対象とするサブグループの違いは、患者レベル分析によって検出されるが、要約レベル分析では検出されなかったことを示した [141]。ある分析では、患者レベルデータの使用により明らかとされた相互作用があるにも拘らず、要約レベルの解析では、治療効果が試験を通して均一であり、メタ回帰分析 (meta-regression) で何の相互作用も特定できなかった。とはいえ、重要な点は要約レベルデータにのみアクセス可能なことが、治療効果の修飾因子 (modifiers) の探索を不可能とするわけではないことである。投与量や投与期間あるいはその他の研究レベルの因子のような研究結果の異質性の潜在的な原因を探ることは、重要なクリニカルクエスチョンを解明する可能性がある。

サブグループ解析に関連するコンセプトは、有害事象のリスク因子を評価するために個別患者データ (IPD) を使用することである。これは、有害事象のリスクの高い患者群を定義する上で非常に価値があり、たとえ、相対リスク尺度で相互作用が認められないとしても、推定された治療効果に絶対的に影響する可能性がある。この絶対的な増加はベネフィット

ト・リスク評価に影響を与える。例えば、特定の非弁膜性心房細動患者のサブグループにおいて、抗凝固剤に関連する出血リスクが高い場合、たとえ相対尺度で治療相互作用が認められるサブグループがなかったとしても、脳卒中のリスク低減とのトレードオフを評価する上で、上記の出血リスクを考慮する必要がある。4.8.1 節では、個別患者データを使用したサブグループ解析について更に議論されている。

全てではないが、状況によっては、幾つかの試験において個別患者データへアクセスすることが可能である。Riley と Steyerberg は、参加者レベルの共変数と事象確率の間にある研究内相関を考慮しつつ、研究横断的に観察される諸リスクの統合を可能とするメタアナリシスのモデルを発表している [69]。このモデルは、個別患者データと要約レベルデータがともに得られるような状況に適用される。彼らは、潜在的な環境バイアスと研究レベルの交絡を説明するために、研究内相関をモデル化し、それらを研究横断的相関から分離することが必要不可欠であると言及している。

3.10 観察研究のメタアナリシスにおける特有の問題

主要事項の要約

- ▶ 個々の観察研究はバイアスのリスクが高く、それらを結合すると悪化する可能性がある。
- ▶ 観察研究のメタアナリシスは以下の 1 つあるいは複数の目的のために検討される可能性がある。
 - ランダム化試験が必要かどうかを決めるため、あるいは利用可能な観察試験の弱点を評価することで、新たな試験のデザインに関する情報を与えるため
 - ランダム化が出来ない介入効果のエビデンス、あるいはランダム化試験で研究される可能性が低いアウトカムのエビデンスを得るため（例：長期的および稀なアウトカム）
 - 慣習的にランダム化試験で研究されない患者グループにおける効果の研究（例：子供、妊婦、高齢の患者）
- ▶ 観察研究による効果推定値をランダム化試験による効果推定値と単一の解析で結合することは推奨されない。
- ▶ 類似しない試験デザインの特徴を持つ観察研究を結合することは、異質性の重要な原因になる可能性があり、それらの結合は慎重にされなければならない。

3.10.1 背景

本書は、医薬品やその他の医学介入の意図しない作用や不都合な作用を確認するという、

困難な領域をターゲットにしている。全てのランダム化されたエビデンスを結合したとしても幾つかの問題について明確な答えが得られない可能性がある。これは、基本的には、有害事象の発生が研究対象とする集団において非常に稀なためである。治験の患者募集は、高リスク患者グループ（例：高齢者、複数の疾患を有する患者、子供、妊娠可能な年齢の女性）が少数であるか、除外される。フォローアップ期間が不十分であり、対象とする集団が小さ過ぎ、あるいは研究されていない可能性がある。多くの治験は、代替（サロゲート）アウトカムを研究する十分な検出力を持つが、臨床アウトカムを研究するには小さい。そのため、幾つかの該当するクエスチョンに答えるために、エビデンスの他の情報源が必要となる[142]。ここでは、観察研究は、通常の治療決定の過程で治療の割り当てを行う介入の効果と安全性（有効性と有害事象）を推定する量的研究と定義される。

観察研究には多くのタイプがあり、それらには、患者間の比較（コホートあるいはケース・コントロール 研究）と患者内の比較（自己対照デザイン）などが含まれる[143]。可能な研究デザインの完全な記述は本書の範疇を超えるが、興味のある読者は、更なる理解のために、疫学と薬剤疫学の一般的な教科書[144, 145].を参照されたい。デザイン評価のフレームワークは Pearce により非常に明確に定められている [146]。

コクラン共同計画は、近年になって、特に彼らの Adverse Effects Methods Group (AEMG) を通じて、観察データが重要である場合があると認めている[87, 147]。経験的なエビデンスでは、有害事象を調べた場合、ランダム化データと観察データのメタアナリシスで類似の結果を示した[148]。以前の研究では、観察研究は相対効果の推定値をより低く見積もることが示されている[149]。観察研究のデザインでは、しばしば、対象とする事象の既知のハイリスク患者、特にその事象の既往歴がある患者は除外されるが、実際の医療では、そのような患者も十分に治療される可能性がある。これは、副作用の絶対発現率が過小評価されている可能性があること意味している。本節では、いくつかのより技術的なあるいは統計的な面を取り扱うが、まずは幾つかの原則的な点に焦点を当てる。

3.10.2 原則

ランダム化されたエビデンスには種々のバイアスがあるが、ランダム化されていない試験では、その全てに加えて非常に多くの他のバイアスもあることは疑う余地がない。主なバイアスは、ある薬剤で治療された患者はその治療を受けていない患者とは大きく異なり、また他の薬剤を投与された患者とも大きく異なる可能性が高い事実と関係している。適応症による交絡（チャネリングバイアスの一形態）は、処方理由が治療と対象とするアウトカムの両方と関連している際に起こる可能性がある。例えば、プロトンポンプ阻害剤は、上部消化管出血の管理に使用され、また胸やけにも使用される。ランダム化された試験は、

この適応症に対するこれら薬剤の効果を提示する。プロトンポンプ阻害剤の使用と上部消化管出血の関係を研究する中では、薬剤曝露に関連したリスクが明らかに逆説的に増加することがある。これは、出血の診断前に発現した症状の治療のために使用された薬剤の頻度の増加に起因する可能性がある。特定薬剤の曝露が、ランダムな割り付けでなく、該当するアウトカムに関連する可能性のある状況によって決定されることから、観察された薬剤とアウトカムの関連は、意図的に過大評価されたり、過小評価されたりする可能性がある。例えば、医師が既存の治療に失敗した患者や重症患者に新規の治療を実施すると、新規治療で有害事象がより高頻度となり、有害事象が治療と関係ない場合は、アウトカムの悪化を招く可能性が増加する。治療群と比較群の相違は完全には解明できない可能性がある。たとえ相違の存在を知ることが可能であっても、相違の調整を完璧に実施することは不可能である。

観察研究は、デザインと実施方法の差異（バリエーション）によって方法論的多様性の可能性がランダム化比較試験（RCT）より高い。観察研究特有の懸念事項として、異なる介入群における患者の違い（選択バイアス）やプロトコルや対象とする特定のクエスチョンについての分析計画を明示的に報告しない研究（報告バイアス）が挙げられる。プロトコルを開示しない問題点の一つは、多くの解析結果から選択的に報告する可能性がある。観察研究の結果にはバイアスの可能性が常にあり、そのため、それらの体系的なレビューやメタアナリシスでは、個別の試験における見込まれるバイアスの方向性や大きさを考慮しなければならない。バイアスのリスクは個別の観察研究で既に高く、個別の研究を結合する際はさらにリスクが高まる可能性がある。疫学研究を評価する際に利用可能な評価ツールの最近のレビュー[150]では、ツールの多くは、薬剤の安全性研究に特に関連する決定的な評価指標を含んでいないことが示された。観察研究のメタアナリシスでは、(a) デザインの弱点、(b) 個別試験の実施におけるバイアスのリスクの慎重な評価、(c) アウトカムの選択的報告を含む報告バイアスの可能性を考慮することが重要である。コクラン共同計画に携わる人たちが開発したバイアスリスクの評価ツールがあり、これはランダム化試験の類似したツールを反映している(<https://sites.google.com/site/riskofbiastool/>)。

上記の制限があるにも拘らず、観察研究のメタアナリシスは、以下の1つあるいは複数の目的のために検討される可能性がある。

- ▶ ランダム化試験が必要かどうか決めるため、あるいは利用可能な観察研究の弱点を評価することで、新たな試験のデザインに関する情報を与えるため
- ▶ 例えば倫理的観点から、ランダム化すべきでない介入についての効果のエビデンス、あるいは長期または稀なアウトカム等、ランダム化試験で研究されることがあり得ないアウトカムのエビデンスを得るため

- ▶ 倫理的配慮などから、慣習的にランダム化試験で研究されない患者群（例：子供、妊婦、高齢の患者）における効果の研究

このことの明確な結論は、ランダム化のエビデンスと非ランダム化のエビデンスを一つの解析で同等（交換可能）なものとして取り扱うのは間違いであり、データの結合ではそれらを区別しなければならないということである。特にベイズの方法論 (Bayesian paradigm) において、両方のソースからのエビデンスを一つの解析において結合するための努力がなされたが、その実施方法は未だ広範には合意されていない。

主要なメッセージは、治療効果の不確実性は標本分散 (Sampling variation) のみでなく、多くの他の要因にも依存するということである。事実、観察データには膨大な患者と事象があることから、標本分散は非常に小さい可能性があり、標準的信頼区間は、ランダム化試験と同様、効果の不確実性をカバーしていない。

3.10.3 観察研究データの妥当性に対する問題

観察研究データの妥当性の問題を検討することは重要である。この問題は、ランダム化試験より観察研究でより一層重要である。

どのようにして研究が選択されるか？

この課題は、明らかにランダム化試験にも適用されるが、観察研究では、より公表バイアスの入り込むチャンスが大きくなる。医薬品の開発初期における多くの治験は、薬剤のスポンサーや規制当局のコントロール下にあり、彼らによく認知されている。もしそうでなかったとしても、多くの場合、特に将来においては、全てのランダム化試験は、プロトコルと共に [Clinicaltrials.gov](https://clinicaltrials.gov) や類似の薬事規制に関連する登録機関へ登録することになる。対照的に、観察研究は関係する薬剤のスポンサーあるいは規制当局にコントロールされておらず、認知もされず、市販後のある時点で実施される可能性がある。観察研究のプロトコル登録制度は、特に公的にアクセス可能な **European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP)** が管理している **E-Register of Studies** [151] を利用した登録が増えており、ある企業団体は事前に定義された項目がある際は、観察研究の登録を推奨しているが [152]、観察研究の中にはプロトコルが登録されない可能性もある。この状況の結果、バイアスの無い試験の選択は非常に難しいものとなる可能性がある。将来のメタアナリシスに含めるために前向きな仕様 (prospective specification) を持つことが可能となる観察研究はめったにない。

多くの場合、観察研究を含めることを検討するときには、その結果を知る事になることから、バイアスの可能性は増大する。大きなサイズの観察研究では、研究者は多くのサブグ

ループ解析を実施する誘惑にかられる。ランダム化試験では、サブグループの解釈は困難であるが、観察研究では、事前に特定していない莫大な数の解析の可能性が大いにある。これらの研究を、特に研究間のサブグループを組み合わせることは問題となる。

各研究において交絡は適切に処置されているか？

ランダム化試験のメタアナリシスにおいて、バイアスの可能性の評価は必要であり、コクランの「バイアスのリスクツール」 **the Cochrane "Risk of Bias" tool** はこれを実施する一つの手段である[62]。観察研究における似たようなツールの開発について既に上記したが、特にチャネリングバイアスの一つである「治療目的による交絡」の面から非常に慎重に評価する必要がある。そのような評価は主観的であり、その結果は注意深く取り扱われなければならない。観察的疫学研究の質改善 **The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE)** ガイドラインが観察研究の報告が適切であったことを保証するのに利用される可能性がある[153]。通常、治療群と比較群の分布を示すだけでなく、該当するアウトカムとの関係性も示している重要なリスク因子の表があるべきである。

研究はバイアスのリスクが低いと思われるか？

ランダム化比較試験(RCTs)のバイアスリスクは多様であるが、観察研究のバイアスリスクの範囲はさらにずっと大きい。メタアナリシスに包含する試験の適合性の判断は、その研究にどの程度バイアスの可能性があるのかによる。既に述べたように、疫学試験を評価するために利用可能な評価ツールの最近のレビューによると、大部分のツールは薬剤の安全性研究と特に関連のある重要な評価要素を含んでいなかった[150]。多くの場合、そのデータは、例えば日常の電子診療記録のように他の目的のために収集されたものであるため、用語の定義が標準化されていない可能性がある。用語を分類するために標準化された手法を用いることと、それを用いたデータが明確に定義されることは同義ではない。多くの欠落データがある可能性があり、これを許容するために使用される手法は慎重に吟味する必要があるだろう。完全な症例分析 (case analysis) においても、必ずしもバイアスが無いとは限らない。

これらの制限はアウトカムの変数とリスク因子の共変数に適用される可能性がある。曝露の定義と信頼性には類似した制限があり、多くの場合、これらの定義は、処方データに依存し、調剤データが利用可能な場合でも、これは参加者が実際に薬剤に曝露されたことを意味しない。時間関連バイアス (Time-related bias) があり、高品質の研究ではこれらのバイアスの可能性を探索し、最小化を試みている。

どの程度一貫性があるか？

研究の異質性を検討することは不可欠である。ばらつきの理由を調査することは、単純に有意差テストを実施するよりも重要である。

研究横断的な一貫したバイアスが結果を説明できるか？

これは多くの考察が必要である。観察研究でベネフィットや他の意図する効果を調査する場合に、より関連する可能性が高くなる。意図しない効果の研究では、一貫したバイアスの可能性は低いことが予想されるが、その可能性を無視することは出来ない。

個別データは利用可能か？

共同研究により個別データが利用可能となる場合がある。重要な事例として、乳がんにおけるホルモン因子の共同研究があげられる。2つの主要なトピックが研究され、ホルモンによる避妊とホルモン代替療法について報告した2つの文献がある。[34, 154]。

これらの研究は、関連データが収集されているにも関わらず主要な研究報告で必ずしも報告が必要とされていない因子に関連する効果を見ることが出来るという面で非常に有用であった。例えば更年期の年齢が一例であり、これは乳がんのリスクへの影響という面で年齢そのものよりも重要であると思われる。

3.10.4 観察研究を用いたメタアナリシスの結論

もしランダム化されたデータが着目した明確な論点に答えるために存在するならば、そのデータに頼ることが最も良い。我々は、一つの解析で、観察研究による効果の推定値をランダム化研究によるそれと統合する事を勧めない。さらに、一般論として、試験デザインの特徴が類似しない試験を統合することは、異質性 (heterogeneity) の重要な原因となる可能性があり、それらを統合する際には慎重に実施しなければならない。これらのデザインの特徴はサブグループ分析や感度分析のよい候補となる可能性がある。

メタアナリシスを用いて観察研究を統合する際は、通常、調整されていない効果の推定値よりも、調整された効果の推定値を解析（すなわち、可能なら交絡のコントロールを試みる解析）するのが適切である。これにより、著者は、一つの試験から報告された代替の調整された推定値から選ぶ必要がある。それでも、調整された効果の推定値の使用により、全ての関連する交絡要素の情報がない幾つかの試験で遭遇する残余交絡の問題が克服されるわけではない。残余交絡の起源は、試験間で異なる可能性が高く、実際には推定値を異なる交絡のサブセットに対して調整することになる。

効果が非常に稀であれば、メタアナリシスは有益である可能性がある。しかし、ランダム

化されていない試験での推定値における不確実性とばらつきが普通の信頼区間では完全には捉えることが出来ないという問題は簡単には克服できない。

関連する問題として、自発的な市販後の副作用が疑われる報告（例：有害事象報告システム）のメタアナリシスは避けるべきである。自発的な市販後報告の最も有用な役割は、着目した安全性の疑問を定義することである。これらの解析は、全体の効果の定量的な評価を含めるのは適切ではない。

最後に、観察試験とそれらの系統立った評価は薬剤の安全性検討において重要であるが、それらには高等な方法論の知識と慎重な判断が求められる。

3.11 ネットワークメタアナリシス

治療法の有効性について比較が必要な際に、直接、比較した試験が全く無いか、あっても少数である場合が多い。しかし、それぞれの治療法は、幾つかの試験でプラセボと比較されている可能性があるため、間接的に比較することが可能である。これを実行するための前提として、間接比較は直接比較より注意深く実施されることが推奨される。新しい治療法が古い治療法より優れていると予想される場合には、いわゆる新規性バイアスが発生する可能性がある [155]。さらに、ダイレクトメタアナリシスに関連する本書の全ての注意事項がネットワークメタアナリシスに適用され、より困難な状況となる可能性がある。

幾つかの問題は、積極的治療を相互に比較する必要性に関連している。図 3.3 に示すように、対象とする比較は、間接比較（プラセボ）を通してのみ推定可能である。

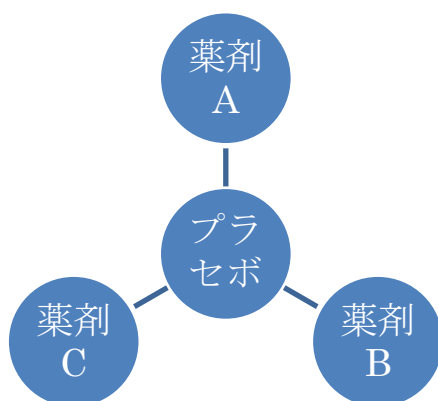


図 3.3 間接比較メタアナリシスが可能なランダム化比較試験のネットワーク (CIOMS X による作成 2016)

図 3.3 では 3 種類の薬剤それぞれが臨床試験でプラセボと比較されているが、どの薬剤も直接には相互に比較されていない。これは、少なくとも、A-P, B-P, C-P の 3 つの試験があることが前提とされる。メタアナリシスでは、それぞれの比較毎にいくつもの試験が存在しうる。この状況では、もし解析の前提が満たされれば、全ての薬剤がプラセボを対照としているという事実のおかげで、ランダム化を維持しながら、それぞれの薬剤を他の薬剤と比較できる可能性がある。

この状況における解析の基本規則は、理解が容易である [156]。もっともシンプルな形として、薬剤 A と B の比較では、試験を統計的に独立であると仮定し、薬剤 A 対プラセボ及び薬剤 B 対プラセボのオッズ比を取り、それらの比を取ることが基本となる。もし二つの試験が統計的に独立しているならば、このシンプルな状況では、信頼区間を推定する事も容易である。

このようなアプローチで有効な結果を得るためには、幾つかの前提が必要となる。どのようなメタアナリシスに対しても、治療間の一対 (pair-wise) 比較に対する効果の異質性 (heterogeneity) の評価は不可欠である。満たされるべき最も挑戦的な前提は、試験デザインと母集団が特定の方法で類似する事である。治療効果に影響する因子、例えば病気の重症度の分布が試験横断的に異なっている場合には、比較は妥当性がない可能性がある。比較している薬剤の治療効果が、軽症患者より、重症患者で大きいと知られている (解析前には知られていない可能性がある) 場合にこのことが当てはまる。重症患者を多く登録する薬剤 A の試験では、軽症患者を多く登録する薬剤 A とプラセボの比較試験より、大きい治療効果の推定値が得られる。薬剤 A の両試験での効果の違いは、当然、病気の重症度によると考えられる。仮に 2 番目の試験が薬剤 B とプラセボの比較を含む場合、特に病気の重症度による治療効果の変化を予想していない場合、効果の違いが薬剤の違いによるか病気の重症度の違いによるかが明確にならない可能性がある。このような役割を担う因子は、必ずしも集団の特性ではなく、試験デザインの要素である可能性がある (例: 盲検下と非盲検化における主観的なアウトカム指標の評価)。なお、治療効果に変化を与えない参加者特性の違いは、このような問題を起こす事がないということも注目に値する。

次の状況は、少なくとも幾つかの直接比較、例えば A-B と B-C がある場合である。繰り返しになるが、これらを比較するいくつもの試験がある可能性がある。そして、複数の試験があることにより、直接比較と間接比較の両方が許容される。これは、「複数治療比較」 (Mixed treatment comparison) と呼ばれる。そして、異なる比較が相互に一貫しているか、つまり同じ治療間での直接比較と間接比較の結果が類似するかという疑問がある。比較の非一貫性や「ばらつき」に関する一般的な問題の解決方法や提言が記載されている複数

の文献が存在する[157-160]。

利用可能なデータが試験結果の要約のみの場合、あるいは個別の研究参加者データがある場合がある。最近の事例として、利用可能な個別参加者データを用いて、非ステロイド性抗炎症薬（NSAIDs）の消化器と心血管への影響を研究している[161]。この論文は、NSAIDsによるこれら二つの有害事象をみているが、個別あるいは要約データのどちらか利用可能な方に基づき、単純に個々の試験に Peto を用い、逆分散法にて直接比較と間接比較における効果を平均した。全体としてあるいは特定の比較に対するネットワークにおける非一貫性は推定されなかったが、それぞれの比較結果の異質性の推定値は得られていた。このような解析を行う際にベイズ法は良く使用され、新しい手法の発見が非常に活発である。

ネットワークにおける観察データの使用について、複数の試みが行われている。その手法の複雑さや、直接比較と異なり、間接比較はランダム化比較とはならず、解析を行う上で最適な方法を示した一般的なガイダンスはまだ無い。特に、データの多くが間接比較である場合や、研究自体が観察研究である際は、解釈を慎重に行われなければならない。

非技術的レベルでこの問題の多くをカバーした比較的、最近の論文がある[162]。さらに、コクランはネットワークメタアナリシスの方法論を適用するために多くの試みを実施しており、以下の URL にて有益な材料を提供している (<http://cmimg.cochrane.org/>)。

第4章 解析とレポート作成

本章の目的は、データ解析の技術的側面に関する情報、安全性データのメタアナリシスに付随する解析及びレポート作成に関するガイダンスを提供することである。技術面に関する広範囲の検討は本章の範疇ではないが、それらに興味がある場合は別の情報源から詳細な情報を入手可能である。しかしながら、解析選択の実用的な意味と重要性については記載している。本章には解釈法も含まれるが、主に技術的な内容を記載する。

本章の主な内容は、治療効果の測定に関する課題、稀な事象の取り扱いに関する課題、適切な統計モデル選択に関する課題、ベイジアンアプローチの長所と短所、検出力の計算と欠測値に関する課題、多重性・異質性・メタ回帰分析および感度分析に関する課題である。

4.1 治療効果の測定

主要事項の要約

- 選択したアウトカムの測定方法は、統計学的有意性あるいは主要な結果の不確実性のレベル、あるいは統計学的異質性の見かけ上の程度に影響し得る。
- データの特異なタイプに応じて、効果と安全性について記述するさまざまな効果推定値を用いることができる。
- 包括的な治療効果に対する尺度の選択において、研究横断的な効果の一貫性、数学的な特性、および解釈が容易であることは重要である。絶対的な測定値（例えばリスク差）は、公衆衛生の介入の影響を相対的な測定値に比べてより良く捉える。一方、相対的な測定値にも長所がある。例えば、実証的研究は、同じメタアナリシスの研究間と比べてより少ない統計的異質性を示すことが分かっている。絶対的及び相対的尺度の双方についての推定効果を表現することはしばしば有用である。

データのタイプに応じて、(有効性あるいは安全性のいずれを扱うにせよ) 様々な効果推定値を利用できる。点推定値とこれらの効果の信頼区間は典型的メタアナリシスの骨格を成している。以下の段落では研究横断的な効果の一貫性、数学的特性と潜在的な測定指標の解釈の容易さを論じる(Deeks & Altman および Sutton らのいくつかの種類のデータについての提言を受けて)。

連続データ：典型的には平均と標準偏差で要約される量的測定値は、高血圧の試験におけ

る血圧、あるいは糖尿病の臨床試験における全ヘモグロビン量中の HbA1c のパーセント (%) のように、しばしば代謝疾患の臨床試験に関する目的の主要あるいは副次アウトカムとなる。このような状況下では、薬剤治療のある特定の用量に対する効果は、しばしば、アウトカムのベースラインからある特定の時点での、実験的治療と対照群間における変化の差 (あるいは変化の%) として定義される。連続変数も安全性の指標になるであろう (例えば、血圧は副鼻腔うっ血 (sinus congestion) に対する薬剤によって上昇しうる)。連続データは多数の統計的手法が使えるため、最も解析しやすい。

それらは標本平均あるいは平均差のようなよく知られた統計量とともにしばしば要約され得るため、これらは通常解釈が容易である。

2 値(Binary)データ: 2 値アウトカムは 2 つのカテゴリーに分類される (例えば、Yes と no、あるいは反応と無反応)。結果は通常、割合として要約される (その事象に関するリスクのある参加者の人数によって、その事象を伴う人数が除される)、一方、治療効果の測定値は通常はリスク差 (2 つの割合の間の差)、リスク比 (一方の他方に対する比の割合)、あるいはオッズ比 (一方の“オッズ”をもう一方のもので除した比、ある事象の“オッズ”とは、割合 / (1-割合) の値である) として与えられる。リスク差 (RD) は絶対的測定値であり、一方リスク比 (RR) とオッズ比 (OR) は相対的測定値である。(本書では“相対リスク”という用語を RR、OR、ハザード比、率比 (rate ratio) などを含む、より一般的な概念を示すものとして用いる。) 定義は用語集(補遺 D)に示した。典型的には、研究者は個々の試験内で、個々の介入群における事象の割合を取り扱う。しかしながら、直接的に治療効果測定値 (たとえば OR, RR あるいは RD) を扱うことが可能となる統計的手法も利用できる。このような測定値の直接的な使用にも、各研究内でのこれらの変動性 (信頼区間あるいは標準誤差) の推定が求められる。

稀な事象に対しては、RR と OR は互いに数値的にかなり近似するものである。研究のいずれの群にも事象の発現がない時、ベイズ法ではこれらのデータをいくらか用いることができるものの、相対的なアウトカム測定値からはその事象の相対的な割合 (rates) についての情報は得られない。効果測定値として RR あるいは OR を用いたメタアナリシスには、双方にではないがいずれかの群に 0 事象の研究が含まれる。両群に 0 事象のある研究は RR あるいは OR に関する情報を提供せず、しばしば数学的にメタアナリシスから脱落する。このように RD の要約は、状況次第では、片方あるいは双方の群において 0 事象の研究も含むことから、相対リスクの要約よりもより多くの研究に基づく可能性がある。

実証的な研究は、同一のメタアナリシスにおける研究間において、相対的測定値の統計的異質性は RD 測定値よりも少ないことを示して来た[165,166]—すなわち、相対的測定値は RD よりも研究間における変動がより少ない傾向がある。しかしながら、RD は一般的には

より臨床的な解釈が可能とみなされている。これらは過剰な数の事象の計算を導くため、相対的測定値よりもよりよく公衆衛生的な影響を反映している。IPDの入手の可能性と全試験の効果サイズの恒常性を推定する意欲に依存することから、一つより多くの指標（オッズ比とリスク差など）を用いた解析が実施される可能性はあり、あるいは一つの尺度（たとえばOR）で解析が行われ、それから要約測定値がもう一つの指標に変更される可能性はある。

LocalioらはORから他の指標へ変換する手法を提案している[167]。この種のアプローチは公衆衛生の理解とメタアナリシスの結果のベネフィット・リスクの推測を助ける点において有用である可能性がある。2%のRDの解釈は3%から1%を引いたものと60%から58%を引いたものでは異なるであろうし、そのためベースラインのリスクと紐付いている。同様に、相対リスクの3は、その事象の率が非曝露患者の1万人中1事象と比較して曝露患者の1万人中3である場合、あるいは非曝露患者の10%に比較して曝露リスクが30%の場合に観察される。これら2つの状況には非常に異なる意味合いが含まれる(5.5.1項参照)。

RRに対してではなくORとRDに対して、結果は事象と無事象の符号化の変更下において不変量のままであることは指摘されるべきである。つまり、事象と無事象の符号化の変更は、効果量、その重要性および観察された異質性へ影響するため、RRの実質的な差をもたらさう。(事象対無事象の選択時期の検討についてのDeeks[165]を参照のこと。)これはRD(これに対しては符号化の変更は効果の徴候のみを変える)あるいはOR(それに対して符号化の変更は本来のORの逆数という効果を生じる)については当てはまらない。

率：カウントデータは発生率として要約されることがある(人・時間あたりの事象数)。この場合、治療効果の測定値は、上記RDとRRの類似の方法において、通常、率差あるいは率比(rate ratio)のいずれかである。

ベネフィットに対する治療必要数(NNT)と害に対するNNT：メタアナリシスにおいて一般的に用いられてはいないものの、ベネフィットに対する治療必要数(NNT)は、投与期間を通じた2値(Binary)反応データにおいて、新たな治療のベネフィットを表す一つの方法である。これはシンプルに2つの比の差の逆数である(つまり、2つの治療グループのそれぞれに注目しているアウトカムの割合)。臨床試験の状況では、大抵は一つの追加のアウトカムを防ぐ／促進するため、ある投与期間を通じて治療を要する患者の数、として説明されう。害に対する治療必要例数の数はよりシンプルな考えであるが、NNTは概して、ある投与期間に一つの追加の有害事象を誘発する治療必要数である。例えメタアナリシスの一部でない場合も、特に信頼区間の計算を重視する場合、NNTを用いることに統計的問題がある[168]。よって、NNTを直接的に使用して研究を統合させることは推奨

されない。他のアイデアとしてはメタアナリシスから単一に統合された RD を取り、それから NNT を算出するというものである。しかしながら、このような統合された NNT は、“真”の NNT が異なる患者サブグループにおいては実質的に変化する可能性があり、臨床家にとってあまり有用ではないであろう。相対リスクの減少が適度に一定にメタアナリシス（あるいは試験のサブグループ）における全試験にある場合、統合された相対リスクの減少（メタアナリシスにおける全試験からあるいは試験のサブグループから計算された）をベースラインの特定の患者群に関連するリスクに応用することは、一連の NNT を計算するために有用で有益である。Smeeth らのこの計算例を参照のこと[169]。

時間事象データ: ハザード比は通常、時間事象データを伴うグループ比較についての主要な効果測定値として役立つ。公表された結果に基づく公表データに依存する場合、時間事象データのメタアナリシスは難易度が高いため、ハザード比は必ずしも報告されとは限らない。それゆえ、ハザード比の算出は、個別の被験者レベルのデータに依存する。種々の変換と近似は、しばしば公表された研究において報告された他の要約統計量を扱うため [170]、また生存曲線からの情報抽出のため [171,172] 提案されてきた。デジタル化された生存曲線も公表論文からのデータ抽出に利用できる。同様に、ハザード比が報告されていない研究がある場合には、エンドポイントを 2 変数として扱う衝動はあり得る。指標の選択はその試験をどう重み付けるかに影響を与え得るということは指摘すべきである。事象発生率が高い時はハザード比の推定値の精度は高いが、OR の予測値の精度は低い。OR を用いることは、その研究の重み付けを実質的にハザード比と比較して軽くし得る（メタアナリシスが逆分散による試験に重みづける方法を用いた場合に）。

序数アウトカムデータ 序数アウトカムデータは、患っている頭痛の程度を反映する患者の自己申告序数スコアのように、神経科学の治療法の臨床試験に時に見られることがあるものの、序数データのメタアナリシスは 2 値 (Binary) や連続データ [173] のメタアナリシスほど一般的ではなく、求められるデータを利用できることは稀である [174]。

上記で検討されたアウトカムの測定値は要約レベルと IPD メタアナリシスの両方で用い得るが、IPD が利用できる場合には明らかに柔軟性がある。同様に、ある連続変数は、二分される可能性や 2 値 (Binary) アウトカム（あるいは率）として分析される可能性がある。例えば、コレステロールを低下させる薬剤の研究において、研究終了時の LDL-C の測定は、その値がある目標レベル(100mg/dL など)よりも低いかなにかによって“Yes”か“No”に二分できる。生存データ（すなわち、アウトカムの変数が注目する事象発現までの時間であるようなデータ）は、全ての研究の期間が類似する場合には、時間事象あるいは OR として解析される可能性がある。アウトカムと生物学的機序の本質は RR, rate ratio ハザード比の選択に影響を与える可能性がある要約レベルのメタアナリシスについては、

個々の研究に対する効果推定値が元の著者による解析に基づき可能な場合には、課題は数種の効果測定値およびそれぞれ異なる推定方法を選択しうるということである。IPD が利用できれば、問われてきたクリニカルクエスチョンに回答するのに最適などのような指標でも用いることで、分析者は一般的に効果量の推定にかなりの柔軟性を持つ。コクラン・ハンドブックの 9.4.4.4 項では 2 値 (Binary) アウトカムに対する測定値選択の検討ポイントについて一般的助言が示されている。

被験者レベルのメタアナリシスについては、包含される全ての研究を通じて個々の被験者について同様のアウトカムの測定値が一貫して集計されるが、効果測定値の選択には依然、効果測定地の臨床上の解釈のような検討を要すると思われる。臨床的な検討に加え、統計的な検討もある。選択されたアウトカム測定値は主要な結果の統計的有意性あるいは異質性の見かけ上の程度に影響する可能性がある。このことから統計的および臨床的な指針に基づく計画の必要性が強調される。到達する結論の頑健性を評価するために感度分析と併せて計画された基本的な準備をすべきである。

いくつかのアウトカムの測定値のタイプ（例えば患者報告のアウトカムのような）については、複合的な評価ツール（尺度）が利用できるであろう。このような状況では、標準偏差単位（いわゆる“効果量”）に示される、標準化平均値差が必要であろう。これらの標準化された平均値差を使う手法については Hedges & Olkin[175] を参照のこと。

4.2 希少事象のメタアナリシスのための統計手法

主要事項の要約

- ▶ 少ない事象数は特有の課題を提示する。これらに含まれるのは：
 - ある研究の一方あるいは双方の群で発現 0 の事象がある場合、関連する指標（例えば log OR と log RR など）はそれらの分散が定義されるため、研究内には定義されない。このためこれらの研究の情報を他の研究に統合することは困難になる。
 - メタアナリシスの標準推定は大きなサンプルの近似に依存する。これらは有害事象の総数が少ないかあるいは有害事象の数がどの単一の研究内においても少ない場合、統計的に有効ではない可能性がある。
- ▶ 少数事象のメタアナリシスにおいて、ある種の手法はバイアスと信頼区間の範囲について十分機能しないため、方法の選択は重要である。
 - 逆数分散とダーサイモニアン・レアード法 (DerSimonian and Laird methods) は希少事象の解析には一般的には勧められない。
 - 連続性の補正は単一群における 0 事象の問題に役立ち得るが、定数（通常 0.5）

を全ての 2×2 表中の全てのセルに加える標準的な補正が最低限望ましい [176,177]。

- ベイズ法は希少事象の解析の分析に対する裏づけを示す。
- ▶ 低頻度の事象のメタアナリシスを市販のソフトウェアを用いて行う場合、そのソフトウェアが実際に行うことを理解することは特に重要である。多くのソフトウェアパッケージには0のセルの自動処理が付いている。そのソフトウェアでは準最適法を用いた0のセルの自動処理しかできない場合、他のソフトウェアの使用の検討が賢明であろう。

重要な健康被害に繋がる希少事象の評価は安全性評価においては一般的である。稀な有害事象のデータのメタアナリシスでは特有の課題が示されることがある。本項では低頻度の有害事象の解析に対するいくつかの課題、警告、解決の可能性のある方法について検討する。

メタアナリシスの標準推定は大きなサンプルの近似に依存する。これらは事象の総数が少ないかあるいは事象の数がどの単一の研究内においても少ない場合、統計的に有効ではない可能性がある。メタアナリシスに含まれる研究数が少ないあるいは多くない場合、異質性に対する統計的検定の検出力は一般的に低い。

一方あるいは双方の群で発現0の事象がある場合、 $\log OR$ と $\log RR$ は定義されず、また、それらの分散も定義されないため、いくつかの潜在的な解析方法に問題を引き起こす。この問題に対処するため、連続性の補正要素はしばしばいずれかの群に0事象を含む研究の 2×2 表中の各セルに加えられる。しかしながら、解析指標が相対的な場合に（例えば RR 、 OR あるいはハザード比のように）、ダブル0（あるいはシングル0ですら）の研究をどう取り扱うのかについてのコンセンサスは得られていない。研究のいずれの群にも事象がない場合、相対的なアウトカム測定値からは相対的な事象の発現率についての情報は提供されない。次の数段落ではいくつか異なる手法と考え方を概説する。

希少事象のメタアナリシスに対する様々な統計的手法が最近の数編の文献に述べられている [176-181]。Suttonらはモデル選択、連続性の補正、正確な統計、ベイズ法、ならびに感度分析にかかわる際に重要な問題のいくつかを検討している [164]。

メタアナリシスにおける稀な事象について述べている数ある文献の中のいくつかを下記に示す。

Sweetingら [176、177]は、 OR を統合するための異なるメタアナリシスの方法の成果と、

連続性の補正の使用における強調とを比較している。彼らは、ダブル0研究は外しており、シングル0試験のみに対して連続性の補正を加えている。彼らは“ここに記載する主要なシミュレーション研究に先んじて行われる予備的な調査は、このような連続修正要素を要さないベイズモデルを用いた研究を含んでいるものと除いているものの両方のデータセットの解析によって、全て0事象 (zero total event) の研究が固定効果メタアナリシスに寄与しないことを裏付けた。”と指摘している。彼らは2つの択一的な連続性の補正と同様に定数補正の要素を研究した。彼らはルーチンに用いられている 2 x 2 表の各セルに0.5という固定値を加える標準的な連続性の補正要素は、特に治療群の実質的なサイズが異なる（つまり、ランダム化率が1 : 1から大きく異なる）ときには、あまり成果が出ない可能性があることを見出した。彼らはほぼ全ての状況において、2つの選択的な連続性の補正は定数の補正要素を上回る働きをすると結論付けている。彼らは同様に、連続性の補正の使用にかかわらず逆分散手法は一貫性に乏しいことを見出している。

Bradburnら[179]は、2変数アウトカムの稀な事象についての10のメタアナリシス手法（7つはオッズ比 (OR) の統合に対して、3つはリスク差 (RD) の統合に対して）の成果を評価した。彼らはまたその試験は“見込まれる傾向も事象の重要性のいずれについての情報も提供しない”ことから、ORに対するダブル0研究を除外した。彼らは統計的手法の選択は基礎にある事象の発現率、見込まれる治療効果の大きさと、この研究における治療群と対照群の数のバランスの検討に依存することを示唆している。彼らは逆分散、ダーサイモニアン・レアード法、および0.5の定数の連続性の補正を用いるマンテル・ヘンツェル

(Mantel-Haenszel) OR法において、バイアスは最大であるということを見出した。事象の発現率が1%を下回るとき（研究されるサンプルサイズに対して）、治療群と対照群の研究内でのサイズに実質的な不均衡はなく、治療効果は特別には大きくないという条件の場合、Peto OR法[182]では、最小のバイアスで最も検出力のある推定値と、最良の信頼区間の範囲が得られる。

Bradburnらは、Gleenland & Salvan[183]が示した考慮すべき、8:1かそれ以上でORが2.5を超える2群ならびにORが5前後のマッチドペア試験における、群間の数の不均衡に伴うPeto法に関するバイアスについて言及している。しかしながら、このような想定は無作為化試験のメタアナリシスでは稀にしか遭遇しない。状況が異なれば、0セル補正なしのマンテル・ヘンツェルOR法、ロジスティック回帰、および正確な階層化OR法は、一般的により成果を上げてPeto法よりもバイアスが少ない。彼らはマンテル・ヘンツェルRD法では相対的にバイアスのない治療効果が推定されることを見出した。しかしながら、事象が稀な場合には、どのRD手法でも幅の広すぎる（そして統計的検出力が乏しい）信頼区間を生じた。信頼区間の範囲はもっとも低い事象発現率における100%の範囲にまで退化した。

Tian ら[180]は 0 事象研究を含む RD を算出する固定効果手順の正確な推定を提案した。Cai ら[184]は稀な事象の状況下での、プアソンの無作為効果モデルに基づく無条件のアプローチ (0 事象研究を含むことによる) を提案した。Bennett ら[178]はベイズと頻度論者の事象までの時間解析データに対するメタナリシス手法を比較した。彼らは 3 つの主要な手法を研究した。まず 1 番目は標準的 Cox 比例ハザード法である。2 番目は罰則付き Cox 比例ハザード法(Firth の補正を伴う Hainze & Schemper 法)で、Firth によって初めて用いられ(1993 年)[185]、その後 Cox モデルに Heinze & Schemper が応用したものである[186]。3 番目はベイズの比例ハザード法(広範な事前分布と有益な事前分布を伴う)である。彼らは Firth 補正法が他に調べた手法よりも優れていることを見いだした。

ベイズ法は稀な事象のメタアナリシスに適切に応用され得る。階層モデルの使用により、治療効果推定に対する後の推定を導くために事象を伴う研究からの情報を取り入れながら、0事象の想定における極端さは調節可能である。ベイズメタアナリシスの稀な事象データに対する実際的な課題は、情報のない事前分布がわずかなデータのために収束不足を導く可能性があるということである。脆弱な情報を有する事前分布は、それは治療効果に弱い制限をつけるが、この問題を解くのに向くであろう。例えば、 $\log RR$ 平均0と標準偏差2を伴う無治療の効果に対する事前分布と仮定すると、治療効果に対するRRは平均のRRは1.0という推定を伴って0.02から55の間であるということが95%確実だということ、およその仮定になる。この事前分布は無情報ではあるものの、RRの可能性のある0.02から55の値の広い幅が特定されるという意味で、合理的に脆弱である。更に安全性の領域では間違いなくより顕著であるいくつかの懐疑的な事前分布を考慮すると、2つの2または0.7の異なる標準偏差を伴う2倍のリスクの効果 (つまり $\log RR=0.7$) を仮定できる。これらの2つの事前分布はRR(0.04, 110)に対する95%信頼区間と (0.5, 8.2) にそれぞれ言い換えられる。これらの3つの脆弱な情報を有する事前分布を感度分析として用いることにより、われわれは事前分布の異なる選択に対抗して結果の頑健性を評価できる。

感度分析 (感度分析についての詳細は4.9項参照) は稀な有害事象の想定においては、統計的および非統計的解釈に対する双方の結果に影響する可能性があるため、特に重要である。統計的問題について、統計的手法の選択、測定尺度、ベイズ法が利用できる場合の事前特定、ならびに0事象研究の分析のために選択された連続性の補正要素は全て評価すべきである。

低頻度の事象のメタアナリシスを市販のソフトウェアを使って行う場合、そのソフトウェアが実際に行うことを理解することは非常に重要である。多くのソフトウェアパッケージには0のセルの自動処理が付いており、例えば、より良い統計的手段を提供するため Sweetingら[176,177]により見出された、他に選択できる連続性の補正の利用機会は無い。

そのソフトウェアでは準最適法を用いた0のセルの自動処理しかできない場合、他のソフトウェアの使用の検討が賢明であろう。

4.3 統計モデルの選択

主要事項の要約

- ▶ 固定効果モデルとランダム効果モデルは異なるリサーチクエスチョンを扱う。モデルの選択は研究内のみの変数もしくは研究内と研究間の変数かがリサーチクエスチョンへの回答に要求されるか否かにより決定される。
- ▶ ランダム効果モデルは効果固定モデルの場合よりも個別の研究の推定により均一な重み付けを生じる。
- ▶ 効果固定あるいはランダム効果モデルを使う決定はメタアナリシスのプロトコルにおいて特定されるべきである。主要なモデルにかかわらず、研究横断的に治療効果のいかなる異質性の理由も探索されるべきである。
- ▶ 効果固定あるいはランダム効果モデルの選択は異質性に依存するべきではない。研究間の治療効果の異質性の程度はメタアナリシスにおいて重要な結果であり、臨床の状況下での解釈がなされるべきである。
- ▶ ランダム効果モデルにおいて、研究間の分散予測の手法は重要である。データが希薄でない場合には、ダーサイモニアン・レナード法、あるいは一般的な混合効果モデルにおける変量効果としての研究特異的な治療効果の包括の少なくとも2つが候補である。

統計的モデルの選択とより一般的なメタアナリシスにおける統計的手法の選択はいくつかの要素による影響を受ける可能性がある。これらにはエンドポイントのタイプ、IPDあるいは要約レベルのデータの有用性、ある事象の発現頻度、研究対象母集団ならびに記述すべき統計的推定が含まれる。この項では統計的モデル選択という重要な面を検討する。この選択は治療効果の点推定に影響しうるため、信頼区間の幅と同様に重要であり、従って統計的意義がある。

4.3.1 固定効果モデルとランダム効果モデル

固定効果モデルとランダム効果統計的モデルは、双方ともにより大きな対象に対するメタアナリシスにおいて観察データからの推定に利用できる。これらのメタアナリシスにおける根本的な相違は、治療効果とその統計的モデルに固定効果としてあるいは変量効果として含まれるかどうかに関わる。どちらを用いるかの選択は、それぞれの意義と同様に時に

メタアナリシスにおいて議論となる[187, 280頁]。

固定効果モデルは時折、一般的効果モデルとして参照される（あるいは少なくとも記述される）。すなわち、それらの研究は全て単一の真に基礎となる効果サイズを推定されていると仮定される。この枠組みにおいて、固定効果とランダム効果の間の選択は効果・サイズのパラメータの異質性の問いによって完全に規定される。しかしながら、一般的効果モデルが固定効果モデルの適切性を示唆するということは真であるものの、その逆は真ではない。固定効果モデルは、これらの研究が全て単一の真の基礎にある効果サイズの推定は示唆しない。

モデルの選択はリサーチクエスションと記述されるべき統計的推定の範囲として情報提供されるべきである。各モデルは異なるリサーチクエスションに言及している[187, 38頁, 188]。リサーチクエスションが平均的治療効果の推定が偶然に一致するか否かに関心がある場合、研究内の変動性（よって固定効果メタアナリシス）のみが適切である。この観点より、固定効果メタアナリシスは、研究の仮定的な対象を推定することに関心があるのではなく、むしろ観察研究におけるサンプリング対象者の偶然の結果を測定しようとする[189]。対照的に、リサーチクエスションがある（仮定的な）研究対象における治療効果の推定に関心があり、既存の研究が（仮定した）その研究のランダムサンプルからなる場合、研究内および研究間の変動性（よってランダム効果メタアナリシス）が求められる。

上記の観点から見ると、メタアナリシスの中のこれらの研究は一般的な治療効果を推定するという仮定は、固定効果モデルにおいて求められていない。同様に、一般的治療効果の欠如はランダム効果モデルの必要性を意味しない。

概ね、ランダム効果モデル（研究特異的な治療効果を変量効果として含むモデル）は、包括的治療効果の推定に対してより大きな標準誤差を生じうる。個々の研究結果への重み付けの観点から見た場合、ランダム効果モデルは固定効果モデルよりも個々の研究推定のより均一な重み付けを生じうる。異質性が高い場合、より均一な重み付けは望ましくもないかも知れないため、ランダム効果モデルの結果として大小の研究に対してのより均一な重み付けの適切性を考慮しなければならない。Poole&Greenland[190]は小規模研究と大規模研究が異なる結果を与えるメタアナリシスの例を示している。彼らは、小規模研究は統計的有意な知見（より大きな効果サイズを示唆している）が示されたときのみ公表されるため、公表バイアスよりこのような状況になりうる、と指摘している。このような状況では、これらは全ての小規模試験のバイアスのかかったサブセットを代表している可能性があるため、公表された小規模研究に重み付けを増すのは不適切である。重み付けと固定効果モデルあるいはランダム効果モデルとを関連付ける必要がないことを

強調することは大切である。重み付けの選択は他の考察により情報提供されるはずである。

リサーチクエスチョンを定義する際に考慮すべき要素は、その後固定効果あるいはランダム効果モデルの選択されるのだが、Berlinら[191]により、特に薬剤開発プログラムにおいて論じられている。

規制環境における安全性のメタアナリシスは、潜在的な治療による現在ならびに将来の患者に対する害を規制当局へ報告されるよう実施されている。このことはしばしば既存の研究における害の推定が偶然に一致するか否かに対するリサーチクエスチョンを導く。上記に述べたとおり、このリサーチクエスチョンは固定効果メタアナリシスにより扱われうる。規制環境においてモデルの選択は規制当局と解析実施の前に合意されるべきである。

固定効果あるいはランダム効果のメタアナリシスが行われたかどうかにかかわらず、メタアナリシスは研究間の効果の一貫性の確認を伴うべきであるということは重要である

(4.3.4および4.8参照)。包括的なメタアナリシスの治療効果の推定は、治療効果の結果がいくつかの研究では治療が非常に有害であることを示し、他の研究では有益ですらある一連の研究から生じうるのと同様に、同一の結果を伴う一連の研究からも生じうる。

介入を伴わない結果の統計的異質性は容易に誤った方向に導かれうる。Petoは次のように指摘している[189]。

例えば、前立腺がんの治療としてのジエチルスチルベステロール (DES) の使用を考えてみると、DESは前立腺がんによる死亡リスクを減じるが、心疾患による死亡リスクを増す。疾患特異的な分析なしでは、この治療が患者に対し心筋梗塞による死亡リスクを高める危険性は容易に見逃されてきた可能性がある。

関連する例はFDAの審査官用ガイダンス[192、7.4.1.1]に以下のように述べられている；

例えばある光毒性の研究事例において、いくつかの研究を統合し、光毒性が弱いことを再確認する結果が得られた。その後、個別に各研究を分析したところ、相当な光毒性を示すある一つの研究が見出された。その研究のみが外来患者についての研究であった（つまり、患者が日光に曝露される機会のある唯一の研究であった）。一定の状況においては、リスクの発生は、様々な研究からある幅を持った値として記述することが最適な場合がある。しかしながら、上記の例のように光毒性の研究では、もっとも適切なデータは、想定される使用に該当する状況下で行われた唯一の研究である外来患者の研究から得られた。

研究デザインと研究結果（異質性）の双方の変動は4.8節で相応に議論されている。ここでは、われわれが繰り返し公表文献に見てきた実践とは、固定効果対ランダム効果モデルの選択のガイドとして変動の程度を用いることだという点のみ指摘しておく。すなわち、人々は単なる結果の異質性の存在をランダム効果モデルの使用理由として用いることがある。下記に論じるが、このアプローチは統計的に妥当性に疑問であるのみならず、結果において異質性の原因への理解を発展させることが分析の中心になるべきであるという、より大局的な点を見逃している。別添2に、エリスロポイエチン刺激剤の例で治療効果の異質性の研究を示す。

4.3.2 モデル特定/分析的手法

固定効果あるいはランダム効果モデルを用いた分析に基づく決定は、プロトコルとメタアナリシスに対する統計解析計画（SAP）において明記されるべきである。主要なモデルに関わらず、研究間の治療効果のいかなる異質性についての理由も探索されるべきであり、異質性探索のための分析的方法もSAPに記載されるべきである（4.8-4.10参照）。

分析的手法はエンドポイントのタイプにより影響を受けることがある。規制環境での治療による害のメタアナリシスにおいて、比率（proportions）、発生率、あるいは被験者レベルのデータが利用可能な場合は事象までの時間の分析に対して手法を適切導くため、エンドポイントはしばしば2変数である。頻度はより低いため、臨床検査値のような連続した変数を用いることにより、害は測定可能性である。主要エンドポイントの変数は、適切な治療効果の測定値（例えばOR,RR,罹患率比、罹患率差、ハザード比、あるいは平均差）と共に、プロトコルおよびSAPに明記されるべきである。

パラメトリック（例えばロジスティック回帰）とノンパラメトリック（例えば Peto 法）両方の、要約レベルの2変数のアウトカム測定値の分析に対する多くの適切な手法がある[182]。これらの方法は固定効果あるいはランダム効果メタアナリシスにおいて利用可能である。しばしばダーサイモニアン・レアー法（DerSimonian and Laird methods）[193]は、ランダム効果モデルにおいて、研究特異的な治療効果の推定における研究内分散を予測するのに用いられる。この手法は算出が直接的で容易であるものの、稀な事象の場合あるいは研究間の異質性に関する推定が重要な場合には制限があることが示されてきた（例えば Jacson ら[194]）。あるいは、多くの統計ソフトウェアパッケージ（例えば種々の SAS 手順）に見られる一般線形あるいは非線形の混合効果モデルは、研究間の治療効果の変動の推定に用いられる可能性がある。これらの手法は変動の推定に通常は尤度を用いるあるいは疑似尤度を用いる手法である。治療効果は階層化ベイズモデルにおいて変量効果として併合もされうる。ベイズのメタアナリシスは医学研究において一層広く使用されるよう

になってきており、読者はベイズ法について論じている本著の 4.4 項を参照のこと。

最近になり、ランダム効果モデルのための分散推定の代替法が注目されている。Cornellら[195]は、いわゆる Knapp-Hartung モデルを含む他のモデルでは統計的有意は得られなかったのに対し、ダーサイモニアン・レアドモデルにおいて統計的に有意な結果を得た例を挙げている。もうひとつ最近の公表文献からはシンプルなダーサイモニアン・レアドモデルの結果を Knapp-Hartung モデルへ変換する計算を含む無料の集計表が提供されている。このように、非常にシンプルなアプローチを“利用しやすい”ソフトウェアを介した、より標準的な手順に取り込む可能性がある。

4.3.3 パラメータ推定

本章ではメタアナリシスにおける包括的な治療効果予測の重要な側面の簡潔な議論を意図している。

通常メタアナリシスにおいて、各研究は独立した推定値と呼応する治療効果のアウトカム測定値の変数を提供する。一般に、重み付けされたこれらの研究特異的な推定値の平均値からはメタアナリシスにおける包括的な治療効果の推定値が得られる。上述したとおり、固定効果とランダム効果モデルは一般に異なる研究の重み付けを生じさせる。例えば、固定効果メタアナリシスにおいて、一般的な治療効果の包括的な推定方法は逆分散法である。

この方法に対し、重み付けされた（メタアナリシス）推定値を算出するために使用される重み付けは、個別研究の治療効果の分散の推定値に反比例する。逆分散法を活用するランダム効果モデルにおいては、個別研究の重み付けは 個別研究のその研究の治療効果の分散の推定値と研究間の分散の推定値の和に反比例する。4.3.2 項で検討したとおり、研究間の分散推定には種々の方法がある。これら 2 組の重み付けを綿密にみると、固定効果モデルとはかなり異なり、ランダム効果モデルでは個別研究の推定値に重み付けできる理由が示されている。上述のとおり、この選択についての実用的な意義があるはずである—これは単に学問的な決定ばかりではない。方法の選択は統計的推定に影響するため、選択の事前設計は必須である。研究間の感度分析と効果の異質性の探索はモデルの選択にはかわらず実施すべきである。

逆分散の重み付けは、そしてどのような一連の重み付けも、各研究の推定に対し相対的な重要性の意味を含むことは指摘されるべきである。メタアナリシスに含まれる各研究は同等に重要であると考えられるべきであろう。このような場合に SAP はこれらの研究が同等に重み付けられることを明記してよい。固定/変量選択が適切な変動性を決定する一方、重み

付けの選択では、個々の研究の推定値が包括的な推定値を得られるように統合される方法を決定する。このような重み付け（より普遍的には解析方法）はクリニカルクエスチョンに一致すべきであり、事前に SAP に明記されるべきである。

治療効果の包括的メタ-推定を得るための上記に述べた手法に代わる手法には、マンテル・ヘンツェル、一般化線形および非線形混合モデル、およびベイズ法などがある(4.4 項)。どの研究も類似の治療効果推定値を有す場合、固定効果およびランダム効果推定値は類似する。固定効果およびランダム効果モデルから異なった要約推定が生じる可能性のあるのは、治療効果推定に研究間の相違がかなり存在するときのみである。それゆえ、異質性における一般的課題が探索されるべきである。ファンネルプロット (Funnel plots) は点推定値がランダム効果と固定効果モデル間で異なる理由を描出するのに有用な図式ツールである。要約推定値を得るために用いられる重み付けと個々の治療効果推定値との関連性は、異なる点推定値を導き得る。上記に指摘したとおり、これはランダム効果モデルを用いてより重み付けされる小規模研究がより大きな効果量を持つ場合に起こりうる。

4.3.4 治療効果の統計学的異質性

研究間の治療効果の異質性の程度は、メタアナリシスにおける重要な結果の一つである。固定効果あるいはランダム効果モデルの選択は、異質性あるいは異質性の欠如に依存すべきでないことは強調されるべきである。あるいは、ランダム効果モデルを用いることは、異質性の潜在的な要因を分析する必要性を事前に排除するものではない。

一般的な異質性の確認方法の一つとして、コクランの Q 統計量[197]がある。他の手法には、特にオッズ比 (OR) について、OR の異質性に対するブレスロー・デイ検定 (Breslow-Day test) とゼレンの正確確率検定 (Zelen's exact test) が含まれる。モデルに基づく方法が治療効果の推定に用いられる場合、異質性は研究の交互作用項により治療を介して検定される。異質性の検定は、研究あるいは事象の数が少ない場合には、臨床的に重要な異質性を検出する検出力が低い可能性があり、または、研究あるいは事象数が大きい場合には、臨床的に意義のない異質性に対する検出力をもつ可能性がある[198]。これは、規制措置に関する決定が治療効果の異質性の程度の解釈に依存する可能性があるため、単に理論上の議論ではない。

異質性を定量化する有用な記述量は $I^2 = 100Q-k+1/Q$ であり、ここで Q はコクランの Q 統計量であり、研究の数である。この統計量は研究間の治療効果の全分散のパーセントであり、すなわち異質性によるものと解釈される。相対的測定値であるため、潜在的な特性によって影響され得る。例えば、非常に大規模な研究が存在する状況では、研究内の変動

の平均は小さくなると予想されるため、効果量について研究間の小さな変動がある場合でさえ、大きな値を生じ得ることを示唆している。従って、異質性の程度は実質的な臨床に関連した状況の中で説明されるべきである [198]。フォレストプロット(4.10.1 参照)は研究間の変動の程度の視覚的な検討に有用なツールとなりうる。

4.4 ベイジアンメタアナリシス

主要事項の要約

- ▶ ベイズ法を使用する利点は多くあるが、最も重要な利点の1つは、統一されたフレームワーク下でエビデンスを統合し、複雑な問題を処理できることである。
- ▶ ベイズ法は魅力的だが、実行が計算上複雑である。マルコフ連鎖モンテカルロ法が使用される場合（例えば、マルコフ連鎖の収束、特定の事前分布に対する感度および初期値がすべて精査されなければならない場合など）方法論の多くの側面をチェックするように注意しなければならない。ベイズ法を採用する際、常に適切な実施を確実にするための専門知識を求めることが推奨される。
- ▶ 感度分析は、ベイズ解析において特に重要であり、事前分布と尤度モデルがどれほどの影響力をもつかを確認する観点から用いられる。

Bayes の定理[200]に由来するベイズ統計学は、確率を厳密に事象の発生頻度に基づいているとみなす頻度論的統計 (frequentist statistics) とは対照的で、確率的声明 (probability statements) を信念の度合い (degrees of belief) とみなす統計学の哲学を指す[59]。概念的にベイズシステムはシンプルである。私達人間は客観的情報に基づいて、私達の意見を修正する。事前情報／初期の信念 (Initial beliefs) と、最近の客観的データは、新しく改良された信念 (belief) をもたらす。この方程式の部分の専門用語は、次の通りである：事前分布は事前情報／初期の信念 (Initial beliefs) の確率分布にあたる。尤度関数は、客観的な新しいデータを有する他の仮説の確率である。事後確率は、新たに更新された信念の確率分布である[201]。

種々のタイプの事前分布がある。例えば、事前の情報、誰かの事前の信念または以前の（例えば履歴的な）データを反映するための"情報のある (informative) 事前分布"でも良い。または、事前の情報、"曖昧"または"情報のない non-informative 無情報事前分布"とすることもできる（例えば、利用可能な関連情報がない場合など）。

例として、免疫抑制活性を有する生物学的物質を見ているならば、日和見感染のリスクが増加することを示す根拠が存在する可能性が高いため、事前の根拠(prior evidence)を使用できる。対照的に、同じ適応を持つ、同種・同効の薬剤が過去にない場合、無情報事前

分布が必要となる可能性がある。

事前分布が指定されると、現在のデータは、尤度関数によって適切な統計モデルで表される。次に、事前確率と尤度を組み合わせることによって得られる関心量の事後確率分布は、ベイズ推論の基礎を形成する。

ベイズのパラダイムとメタアナリシスの両方におけるゴールは、可能な限り高い精度によって推定を引き出すこと、推定に関連するすべての不確実性を考慮すること、そして首尾一貫した決定につながる形で結果を提示することである。

4.4.1 ベイズ法の長所と短所

ベイズメタアナリシスは、頻度論的アプローチに比べて多くの利点を提供する。以下に箇条書きとして、主な長所と短所を要約する。

ベイズ法の長所

- 異なるレベルのすべての不確実性を考慮しながら、公式かつ統一された一貫性のある形式で、複数のデータソース/研究/治療からのエビデンスを統合するフレームワークを提供すること。理論上、これを頻度論的フレームワークによって実施することは可能だが、しばしば扱いにくい。したがって、頻度論的な方法は一般に、いくらか簡単なモデルを使用する（例えば、研究間の効果サイズの分散などの不確定なパラメータが、実際に分かっていると仮定するが多い）。マルコフ連鎖モンテカルロ法（MCMC）をベイズ階層モデリング手法と組み合わせて使用することで、潜在的に複雑なメタアナリシスの様々な問題に対処することができる（時間の経過とともに非定常的なハザードを伴うイベントを処理する場合、またはリスクが特定薬剤における特定の適応症にのみに適用可能な場合）。
- 関心のある統計量（quantities of interest）の事前分布を利用することにより、他のエビデンス（ヒストリカルな対照データ、同じ薬効群の他の薬剤データなど）を正式に組み込むことができる。
- 特定の仮説に関する確率の計算を可能にする。例えば、治療効果を表す OR が 1.3 未満である確率（これが高確率であれば、その事象が医薬品によって引き起こされていないという帰無仮説を支持するかもしれない。これは FDA からの抗糖尿病薬開発における心血管リスク評価のための要求を連想させる）[8] または治療効果を表す OR が少なくとも 2 である確率（これが高確率であれば、臨床的に意味のある治療効果の存在を支持するかもしれない）である。このような確率の計算は、パラメータ推定の古典的な頻度論的フレームワークでは実行できない。さらに、P 値や信頼区間などの

古典的推論よりも理解しやすい異なるスケール（OR、RR、RD など）での真の治療効果についての確率を述べる事が可能である。例えば、薬剤 A と薬剤 B との間の死亡率の RD が被験者 1000 人中 1 人より大きい確率や、薬剤 A を服用している被験者が薬剤 B よりも生存期間中央値が長い確率に興味があるかもしれない。

- ▶ 不確実性の予測を含む、予測に関する意見を容易に作成できるようにする。
- ▶ まれなイベントでのメタアナリシスで追加の利点を提供する。これらは 4 章 2 節で議論される。

ベイズ法の短所

- ▶ 関心のあるパラメータである事前分布の仕様は、ベイズ法の中で最も困難で論争の多い側面の 1 つである。査読者は、事前分布の信念が客観性を損なう可能性がある懸念している。この章では、客観的データに基づいて事前分布を注意深く指定するか、弱い（weakly）事前分布（4 章 2 節を参照）を使用することが重要である。規制面では、主観的な信念に基づく優先順位づけは避けなければならない。しかし、真に分布が無情報事前分布であれば、ベイズ法が分布の影響を受けていないことを確認するために必要な、余分な作業を回避する同様の頻度分析が利用可能である可能性がある。これに対する例外は、上述のベイズ法の長所の箇条書き冒頭に記載されているように、ベイズアプローチが計算上優れている複雑な問題がある場合である。

4.4.2 追加情報

頻度論に基づくランダム効果のメタアナリシス手法は、通常は全体的な治療効果の推定に焦点を当てるが、完全なベイジアンアプローチは、全体的な治療効果と試験特有の効果（全体的な平均に縮約する）も推定でき、さまざまな有益な拡張が可能である。膨大な文献によって、多種多様なメタアナリシスの問題[199、202-207]に対処するためにベイジアンランダム効果モデルが中心に発展してきた。この研究のすべては、モデル化の仮定を介して間接的に、またはより直接的に適切な回帰の構造とともに追加情報を導入することによって、異質性の異なる側面を扱うことに焦点を当てている。ベイジアンメタアナリシスの実用的な例がいくつかある。ある研究では、非常に少数の研究によるメタアナリシスは固定効果モデルを“初期設定”にする必要がないという事実について議論されており、これは異質性の程度を情報のある事前分布として、単純なベイジアンアプローチに組み入れられるからである（Higgins et al）。Higgins & Spiegelhalter は、大規模試験の公表の前後に心筋梗塞におけるマグネシウムの治療に対し、ベイジアンメタアナリシスを適用した[209]。彼らは、懐疑主義（scepticism）を事前分布として、公式（形式）的に分析に組み入れる方法を示している。彼らはまた、ベイジアンメタアナリシスのモデルを用いて、治療効果

が試験の規模または対照群の内在するリスクのいずれに依存するか、という異なる仮説を探索する方法を示した。Kalil [49]は、FDA および Yahav らによって頻度論的方法を用いて得られた結果に対して、ベイジアンアプローチに基づく再評価を実施したところ、Cefepime(セフェピム)は好中球減少症の患者の死亡率を、それぞれ 91% (FDA による試験レベルのメタアナリシス)、81% (FDA による個別患者レベルのメタアナリシス) および 99% (Yahav らによるメタアナリシス) 上昇させることが示された。この例は、ベイジアンアプローチに基づくと直接的に確率に言及できることを示している。これは、異なるアプローチが安全性シグナルについて異なる結論を導き出す例でもある。Askling らは、ベイズ階層区分指数生存モデルを用いて、腫瘍壊死因子(TNF)阻害薬クラスの癌リスクを調査した[210]。ベイズモデルは、個々の参加者レベルのメタデータを分析し、参加者レベル、場合によっては時間依存性共変量や試験の異質性を考慮したモデルを作成することが可能であった。Kaizar らは、抗うつ薬を使用する子供の自殺リスクを定量化するためにベイズ階層モデルを用いた[211]。Ibrahim らは、抗糖尿病薬開発の PhaseII/III 試験計画のためのベイジアンメタアナリシスのサンプルサイズ決定法を開発した[212]。

4.4.3 研究の統合における重要な留意点

ベイズ法は魅力的であるが、実行するのが計算上複雑である。

方法論の多くの側面（例えば、MCMC 法が使用される場合、マルコフ連鎖の収束、特定の事前分布に対する感度、および初期値がすべてチェックされるべきである）をチェックする必要がある[199]。ベイズ法が適用されるときは常に、適切な実施を確実にするための専門知識が求められるべきである。

歴史的には、計算時間が問題だったが、最近のソフトウェア開発の進歩がこの状況を改善した (WinBUGS、SAS PROC MCMC、JAGS、特に Stan)。

感度分析は、ベイズ法にとって特に重要である。感度分析の一般的なアドバイスが 4.9 で示されているが、我々はベイズ法にとって特に重要なことを記載する。

任意の与えられたベイズ法は、いくつかの事前分布を有する可能性がある。

事前分布がどのような影響力を持っているかを確認するには、感度分析が必要である。

さらに、他の伝統的な統計手法を使用する場合と同様に、適切な統計モデルを使用した感度解析も実行する必要がある。

4.5 検出力 (Power)

主要事項の要約

- ▶ 規制目的のためのメタアナリシスのほとんどは、分析時に関連するすべての既存のデータを分析するため、試験の検出力は計画目的 (**planning purposes**) ではなく記述目的 (**descriptive purposes**) で使用される。しかし、規制面では、規制当局とスポンサーとの間で前向きに計画されたメタアナリシスが合意される可能性があり、この場合、検出力はプロトコルの一部でなければならない。
- ▶ ランダム効果に基づくメタアナリシスは、一般に、研究間のばらつきを含めるため、固定効果に基づくメタアナリシスよりも検出力が低くなる。
- ▶ 研究間またはサブグループ内、または全体の治療効果の異質性の検定は、全体的な治療効果の検定よりも検出力が低い可能性が高い。
- ▶ 交互作用または複数のサブグループの試験は、臨床試験のように、メタアナリシスにおける検出力およびタイプ I エラーに関して同じ弱点を有する。

本節の目的は、メタアナリシスにおける統計的な検出力に関する重要な考察を示すことである。統計的な検出力の概要または検出力計算の公式を示すことを意図していない。

メタアナリシスの検出力は、ある対立仮説が真である場合に (有意水準 α のもと) 帰無仮説を棄却する確率と定義される。この議論のために、対立仮説は、治療間の真の効果量が θ であると仮定する。検出力は、臨床試験のための試験のサンプルサイズを決定するために使用される重要な計画の概念であるが、前向きメタアナリシスにおいてはほとんどの場合、分析時における全ての既存データを分析することから、あまり重要ではない可能性がある。しかし、規制面では、規制当局とスポンサーとの間で前向きに計画されたメタアナリシスにおいて、検出力はプロトコルの一部に含まれるべきである。

例えば、関心のあるパラメータが **OR** であり、これがすべての研究に共通であると仮定されている場合、背景のイベント発生率および相対リスク θ の仮定について、 θ を検出するための特定の検出力を得るのに必要なイベント数を算出し、前向きメタアナリシスが実行可能かどうかを判断するために使用される。同様に、非劣性の設定では、相対リスクの特定の値を排除するために必要な事象の数も計算することができる。

より現実的には、単一の値 θ は全ての研究に共通するわけではなく、研究間のばらつきを伴った研究間の研究固有の値 θ の分布であってもよい。全ての研究に共通の θ がない場合、全体的な治療効果は、個々の試験の推定値を組み合わせることで全体的な推定値を得る方法によ

って決定される。個々の研究の推定値を組み合わせる方法は、研究間のばらつきを組み込むことができる（しかし、必ずしもそうする必要はない）。

したがって、治療効果の標準誤差に研究間のばらつきを組み込んだランダム効果モデルでは、メタアナリシスを実施する前に検出力を計算するために、 θ_i のばらつきに関する仮定が必要となる。

メタアナリシスでは、検出力に関して留意すべきいくつかの重要な点がある。これらのポイントの一部は、Hedges & と Pigott[213,214]から派生した。

- ▶ ランダム効果に基づくメタアナリシスは、一般的に固定効果に基づくメタアナリシスよりも検出力は劣るが、分析に θ の研究間のばらつきが含まれているためである（この余分な検出力は、第1種の過誤の増加を伴う）。例えば、メタアナリシスは、大きな治療効果をもつが、イベント数が比較的少ない研究を含めることができる。重みが分散の逆数である場合、この研究は、固定効果モデルよりもランダム効果モデルにおいて比較的大きな重みを与えられ、より大きな効果量および場合によってはより大きな検出力をもたらす。要点は、推定されるパラメータは分析モデルの選択に依存するため、メタアナリシスにおいて帰無仮説から分離する能力は、分析方法の選択に依存することである。
- ▶ 試験間の治療効果の異質性に関する検定は、交互作用の検定であり、特に交互作用が逆方向ではなく治療効果の程度の1つである場合、全体的な治療効果の検定に比べて強力ではない可能性が高い。この場合も、必ずしもそうであるわけではない。例えば、全ての研究に対する治療効果はゼロであろうが、研究全体の治療効果のばらつきは大きい可能性がある。
- ▶ 同様に、サブグループ全体での差別的治療効果の検定（研究のタイプ別または研究内の要素別）は、特に交互作用が反対方向ではなく治療効果の程度の1つである場合、全体的な治療効果の検定に比べて強力ではない可能性が高い。このような低下した検出力のため、サブグループ解析における有意性の欠如は、治療効果がサブグループ間で異ならないとする決定的な証拠として解釈されるべきではない。
- ▶ サブグループ分析では、交互作用の効果の大きさが平均的な全体的な治療効果より大きくなり、それによって交互作用の検定のためにより大きな検出力が得られる可能性がある。統計的に有意な交互作用、特に全体的な治療効果がない場合には、探索的発見として解釈されるべきであることを強調する。この交互作用を調べることがメタアナリシスの第一の目的であったとしても、サブグループ内での治療群の比較可能性には依然として懸念がある（第4章4.8.1項参照）。
- ▶ 有意な全体的治療効果がない場合、点推定値および95%CIは、事後的な検出力よりも結果を解釈する上でより良い指標である。信頼区間の上限及び下限に、事後的な検

出力よりも大きな重みを与えることが推奨される。

4.6 欠測値

主要事項の要約

- ▶ メタアナリシスにおいて欠測値を生じる可能性があるのは、研究全体の欠落、対象とするアウトカム（あるいは統計量）の欠落、RCT内の一部個人からのデータの欠落である。
- ▶ すべてのデータアナリストは、欠測値の問題に直面している。この節では、RCTのメタアナリシスで発生する可能性のある欠測値の問題の概要と、それら进行处理するための一般的なアプローチについて説明する。安全性のアウトカム（safety outcomes）特有の検討事項が特定されている。

4.6.1 欠測値の概要

システマティック・レビューやメタアナリシスにおける欠測値の潜在的な原因は、次の3つの主要なカテゴリーに分類することができる。

(i) 研究全体のレビューからの欠落。 (ii) レビュー対象とするアウトカム (outcome interest)の個々の研究からの欠落。 (iii) RCT内の一部個人からのデータの欠落。

(i) 研究全体のレビューからの欠落

研究全体の欠落は、安全性データのメタアナリシスに特有の問題ではない。研究の欠落の一般的な例としては、未発表の研究や外国語で発表された研究が挙げられる全ての研究が公開されているわけではない場合、利用可能なエビデンスによって全体を評価しようとすると、メタアナリシスの妥当性に大きな脅威が生じる可能性がある。未発表の試験が発表された試験よりAEの情報源として優れている可能性があるため、安全性データのメタアナリシスについても同様な問題が生じる[215]。メタアナリシスに関する主な懸念事項は、出版バイアスの可能性、すなわち、研究結果の性質に基づいた選択的出版の可能性である。出版バイアスを最小限に抑えるために、幅広い出版および未発表の研究を特定するのに役立つ包括的な検索戦略 (comprehensive search strategy) を使用すべきである。3.8節で、この点についてさらに詳しく説明する。

Funnel plots [132]は、潜在的な出版バイアスを検出する手段である。3.8.4項には、バイアスがない場合の対称ファンネルプロット (symmetrical funnel plot) とバイアスのある

場合の非対称ファンネルプロット (asymmetrical funnel plot) を含むファンネルプロットの詳細な説明が記載されている。

出版された研究に基づく分析において、出版バイアスは、出版されなかった研究を解析できないことに加えて、研究の方向性や研究結果の統計的な有意性を好む学術誌の傾向に基づいて出版されることから起因する (例[216])。肯定的な結果が含まれる研究は、多くの場合、後続の出版物において引用される可能性が高く、検索戦略が包括的でない場合に研究に影響を与える可能性がある[217]。出版バイアスの問題は、典型的には、有効性解析において生ずることに留意すべきである。しかしながら、関連する安全性データを報告する研究が欠落する可能性も無視してはならない。

研究結果は、出版の回数に影響を与える可能性があり、同じ研究が複数回出版されることがあるため、重複したデータがメタアナリシスには含まれないように注意する必要がある[6],[218]。肯定的な結果の試験は重複して出版される可能性が高まるため、治療効果を過大評価する可能性がある。個々の試験結果にかかわらず、研究仮説に関連するすべての試験を含めることが目的でなければならない。

(ii) レビュー対象とするアウトカムの個々の研究からの欠落

RCT からのアウトカムデータの欠落は、実験的治療の安全性を推定する上でより大きな不確実性を生じるとともにバイアスを生じる可能性がある。アウトカムデータの欠落を生ずる主な理由は 3 つある。

RCT の安全性データのメタアナリシスでアウトカムが欠落する理由の 1 つは、不完全な報告である。多くの研究では、一連のアウトカム測定値を収集しているが、すべてを報告するわけではない。RCT からどの結果を報告するかの決定が、その研究結果により影響されれば、おそらくメタアナリシスの結果のバイアスとなるであろう[219、220]。不完全な報告は、安全性データのメタアナリシスにとって特に問題となる可能性がある。例えば、臨床試験の報告では、通常、有効性の結果が強調され、有害事象の詳細な文書化が行われていないことが多い[126,127,221,222]。場合によっては、有害事象に関する情報を全く提供しない研究もある。また、研究によっては、元の試験担当者 (original investigators) によって薬剤と関連があると考えられる有害事象のみ報告されることがある。頻繁に発生する有害事象のみを報告する例もある (例: 「治療群において 2% 以上発現した事象を報告する」)。安全性データのメタアナリシスの問題の 1 つは、有害事象が多くの理由で欠落している可能性があり、有害事象の不完全な報告の理由を確認するのが困難なことである。最終報告書からデータを除外するルールに矛盾が生じる可能性があるため、データを分析

的に修正しようとする、はるかに大きな問題が生じる可能性がある。

アウトカムデータ欠落の第 2 の理由は、元の研究における報告される要約統計量 (summary statistics) の不一致から生じる。メタアナリストは、元の研究から得られた情報では、メタアナリシスに含めるのに適した形式 (OR、RR、HR など) でアウトカムの要約統計量 (summary statistics) を計算できないことがよくある。例えば、個々の介入群のサンプルサイズ、事象の数、ばらつきの指標 (標準誤差、信頼区間など)、参加者の追跡期間、イベント発現までの時間 (time-to-event outcomes) などが研究の報告から欠落していることがある。

アウトカムデータが欠落している第 3 の理由は、元の研究における不適切な分析から生じる可能性がある。例えば、特定の試験報告書は、治療効果を計算するために使用された欠損データを処理する方法を明確にしていないことがある。メタアナリシスは、様々な分析方法を用いて得られた推定値を組み合わせることになり、メタアナリシスの妥当性を危うくする可能性がある。

一般に、アウトカムデータ欠落の同定は、研究をシステマティック・レビューまたはメタアナリシスから除外する理由として使用すべきではない。可能であれば、メタアナリストは、元の試験担当者に連絡して、最終報告書で明らかにされていない特定のアウトカムに関する欠落した情報、欠測値の解析方法や処理方法に関する情報を求める必要がある。

(iii) RCT 内の一部個人データの欠落 :

ITT (治療意図) 解析が行われない場合 (すなわち、すべてのランダム化された参加者が分析に含まれていない場合)、解析されずに残る参加者データが存在する。他の例には、ノンレスポonder、参加者アウトカムの記録不備、参加者の減少などが含まれる。データ欠落の別の側面は、対象とするアウトカムのリスク因子に関連する情報となることであり、例えば、治験からドロップアウトする参加者の差の情報は、実薬群とプラセボ群との潜在的な不均衡の評価を可能とする。

4.6.2 メタアナリシスにおける欠測値の取り扱いに一般的に用いられる方法

欠測のあるアウトカムデータは、適切に分析されなければ、誤った治療効果の推定につながるため、RCT の結果の妥当性に対する共通の脅威である。RCT における個々のデータの欠測を処理するため、i) LOCF (last observation carried forward) ii) 最小二乗法 iii) 補完法、および iv) 尤度に基づくアプローチなど多くの統計手法が開発されている [223] 。

多重代入法、パターン混合モデル、利用可能なデータの完全尤度分析などの高度な手法においては、各参加者の詳細なデータが必要である。RCTの要約データは多くの場合、公表されたレポートから得られ、RCT参加者の個別データの利用は制限されるため、要約データのメタアナリシスでは高度な統計手法の利用は困難である。

欠測のあるアウトカムデータを含む研究は、この研究を含むメタアナリシスの妥当性を脅かす。コクラン・ハンドブック[6]と英国国立医療技術評価機構(NICE)[224, 225]は、治療効果の推定において、欠測データによって引き起こされる潜在的なバイアスや不確実性を説明するために、メタアナリシスを実施する上で、欠測データをどのように扱うべきかに関する勧告を提供している。これらの勧告は下記を含んでいる。

- 1 個々の試行で欠落したデータによるバイアスのレベルを判断する。
- 2 この潜在的なバイアスの影響を探る。
- 3 不足しているアウトカムについて異なる仮定を置くことによって、多数の感度分析を実施する。

偏りを考慮し、欠損データによって引き起こされる治療効果推定値の不確実性を定量化するために、異なる代用戦略またはベイズ統計手法を使用する統計モデルを使用する方法を示す2つの有用な論文は、Turner et al、Higginsらのものである[226, 227]。

要約すると、RCTのメタアナリシスにおける欠損データは、レビューから欠落した研究、要約データの欠落または不十分、またはRCT内の一部の個人の欠落データの形で生じる。研究全体がレビューから欠落している場合、出版バイアスの特定と出版バイアスに対する結果の感度の点検が考慮されるべきである。RCT内のメタアナリシスへの関心の結果がRCTにない場合、またはRCT内の一部の個人に欠けているデータ問題が発生した場合は、欠落したデータを処理するために前述の方法論を考慮する必要がある。さらに、欠落したデータがレビューの所見に与える潜在的な影響を明確に取り上げるべきである。メタアナリシスが出版された論文の要約情報に基づいている場合、データの欠落量とデータが著者によって処理された方法は、方法論的な個別の試験を評価する際の考慮事項となる。可能であれば、紛失したデータや紛失したデータの取り扱い方法に関する論点に、元の著者に連絡する必要がある。

4.7 メタアナリシスにおける多重性と累積メタアナリシス

主要事項の要約

- ▶ 累積メタアナリシスには、1 つまたは複数の研究からの新しいデータが蓄積されるため、複数のテストの前向きな計画が関与する。
- ▶ タイプ I のエラー率は、メタアナリシスを通じて複数の試験（例えば、複数の用量、アウトカム、サブグループ、期間、効果測定、累積データを見る）により大幅に増加する可能性がある。
- ▶ 主要アウトカム指標があらかじめ定義されていない場合、アウトカムの測定対象の日付駆動型選択により、偽陽性率がさらに増加する可能性がある。
- ▶ 複数のテストによる偽陽性率の増加の問題は、実用的かつ満足のいく解決策がないために、ほとんど無視される。
- ▶ ほとんどのアプローチでは、トライアル数（試験数）や試験ごとのサンプルサイズを事前に知る必要があるが、通常はそうではない。累積メタアナリシスはプログラムの計画中に計画されている場合でも変更の対象となることが多いため（特に適応的試験 [adaptive trials] の場合）、不確実性が存在する。
- ▶ ベイズ法は、一般に、関心のある量に関して直接確率を述べることを可能にする一方で、他の局外パラメータ（nuisance parameter）の不確実性は分析において自動的に考慮される。

メタアナリシスで複数のテスト（例えば、複数用量、アウトカム、サブグループ、期間、効果尺度、累積データの分析などについて）を分析すると、タイプ I のエラー率、安全性（または有効性）に関する誤った肯定的な結論を導く可能性（実際には差異がないにも関わらず、治療上の差があると結論づける）は、大幅に増加する。多重性が生じる 1 つの状況は、累積メタアナリシスの状況にある。累積メタアナリシスは、特定の順序で（例えば、研究完了または公表のデータに従って）1 つずつ研究が追加され、新しい研究が追加されるたびに結果が要約されるメタアナリシスである。累積メタアナリシスは、新しい研究が完了したきにエビデンスの本体に追加されていく前向きに計画された一連のメタアナリシスと考えるのが有益である。累積メタアナリシスには、1 つ以上の研究からの新しいデータが蓄積される際に統計的有意水準をどのように調整するかを明記した正式な計画が含まれている必要がある。複数のエンドポイントが存在する場合、時間に関して同じクエスション（1 つのエンドポイント）を扱うことにフォーカスしている累積メタアナリシスの多重性の問題は拡大する。

Biester&Lange[228]は、Cochrane library から得られたシステマティック・レビューの無作為サンプルを調査した結果、アウトカムの数は 1 から 35 まで変化し、アウトカムの中央値が 6 前後であることを報告した。さらに彼らは、累積的なメタアナリシスの 47% だけが、プライマリーとセカンダリーのアウトカムを区別しようとしていることを指摘した。

主要なアウトカムがあらかじめ定義されていない場合、結果の指標をデータによって選択することにより、偽陽性率がさらに上昇する可能性がある。例えば、進行中の一連の研究では、いくつかの試験が終了した後に、主要な安全性のアウトカムの指標を定義することもある。その主要な安全性のアウトカムの指標の定義は、事後的であり、終了した試験からの結果から導き出されているであろう。ここで得られる結果は、最上の場合でも探索的と捉えるべきであろう。

初期の累積メタアナリシス・ペーパーでは、複数のテストによる偽陽性率の増加という問題はほとんど無視されていた[229-231]。Biester&Langeの調査で明らかのように、累積メタアナリシス論文の著者のわずか4%が多重性調整について何らかの声明を出し、2%が調整を行ったと述べている。多重性は、累積メタアナリシスにおいて一般的に無視されている問題であると結論づけた[228]。

何人かの著者は、偽陽性率を制御する方法を提案している。偽陽性率を制御するための一連のアプローチは、数名の著者によって議論されている[232-234]。Whiteheadは、医薬品開発プログラムの中で治療計画が有効かどうかをできるだけ早く決定するために、一定数の中間解析を前向きに計画するトライアングル境界を用いることを提案した。Pogue&Yusuf[233]は、O'Brient-Fleming型アルファ spending アプローチのような従来のグループ・シーケンシャル法の適用を検討した[235、236]。van der Tweel&Bollenの論文[234]では、これらのアプローチ（著者らによってトライアングル境界のシーケンシャルメタアナリシス、O'Brient-Flemingタイプ境界のトライアルシーケンシャルメタアナリシスと名付けられている）のパフォーマンス特性を比較している。これらのアプローチは、試験の数およびまたは試験内のサンプルサイズを知る必要性から批判されている。累積メタアナリシスはプログラムの計画中に計画されている場合でも（特に適応試験の場合）変更の対象となることが多いため、不確実性が存在する。シーケンシャルアナリシスに関するさらなる開発および議論は、Kulinskaya&Wood、Imbergerら、Miladinovicら、Higginsら、Thorlundら、BrockらおよびWetterslevらの論文で得ることができる[237-243]。

いくつかの臨床開発プログラムでは、メタアナリシスが繰り返し行われ、利用可能なデータが決定しない場合、追加の安全性試験が開始される可能性がある。1つの例は、糖尿病薬に対してFDAが要求した心血管イベント試験である。

Jennison&Turnbull[244]は、これらの種類のプログラムの適応の性質に対処するために、累積メタアナリシスにP値の組み合わせアプローチを適用している。適応的かつ繰り返される安全性の累積メタアナリシスは、多重性の観点からさらに多くの課題を提示する。

Quan ら[245]によって提案された頻繁論的アプローチ（フィッシャーの P 値の組み合わせテストの改訂版）とベイズアプローチは、タイプ I のエラー率の制御に適用可能であり有用である可能性がある。

Lan ら[246]は前向きに計画された、または後ろ向きに調べられた量的エンドポイントの累積メタアナリシスにおいて、複数の検定を考慮するために、統計量の Z 値を不利にする重複対数の法則（LIL: The law of the iterated logarithm）に基づくアプローチを提案している。この方法はまた、試験間の治療効果における異質性の推定も説明する。Hu ら相対リスク、オッズ比、またはリスク差の 2 変数のアウトカムの分析にこの手法を広げている。彼らの手法は柔軟であり、分析は様々な時点で行われ、最大の情報は未特定のままである。しかし、研究間の異質性が合理的に小さく、最大限の情報があらかじめ指定されているシナリオでは、LIL アプローチを用いることにより、大きな検出力の低下が発生する可能性があることに注意する必要がある[247]。例えば、参加者が極端に多様化することが予想されない場合、試験内監視のための LIL アプローチの使用は勧められない。これらのシナリオでは、 α 消費関数（alpha spending function）などの従来のアプローチは、タイプ I エラー率を制御でき、LIL アプローチよりも検出力が高い。LIL アプローチが開発された研究間の異質性が多いまたは未確定の最大情報が多いシナリオでは、一般的に使用されている他の方法が、タイプ I エラー率を防がない可能性があるため、LIL アプローチが他のものと比較してどのくらい控えめであるかは明らかではない。Hu らは、タイプ I エラー率を制御し、任意の検出力を達成するための適切な調整係数を決定するためにシミュレーションを用いることを推奨している。

これらのアプローチは、名目レベルでの偽陽性率の制御に役立つ可能性があるが、一連のアプローチの試験回数やサイズを知る必要がある、また、LIL アプローチの差の検出力をさげるなど、限界がある。さらに、安全性のアウトカムの累積メタアナリシスについては、特に規制の観点から潜在的な安全性のシグナルの検出力が重視されるため、多重性の調整が受け入れられないという議論になる可能性がある。多重性の調整が実施されていない場合、そのような累積メタアナリシスからの潜在的な肯定的な所見の解釈は、偽陽性所見の可能性があるという観点で注意深く評価する必要がある。このような場合、推論ではなく推定に焦点を当てるべきであり、エビデンスは仮説形成と見なされる可能性がある。名目上の P 値（nominal P value）と調整後の P 値が異なる結論を導く可能性があることに注意しなければならない。エンドポイントが死亡率などの非常に重症な有害事象である場合、これらの考察がより重要となる。

Lau ら[248]によると、累積メタアナリシスはベイズフレームワークで最も自然に解釈される。このフレームワークでは、現在のデータを得る前に利用可能な知識は、事前分布に組

み込まれ、現在のデータが処理された後に利用可能な知識は、事後分布によって表される。この事後分布は、より多くのデータが得られると、新しい事前分布になる。ベイズ分析は、事前確率をモデルまたは仮説に割り当てることによる多重性の調整のみを扱う。ベイズ法は、一般に、関心のある量に関して直接確率を述べることを可能にする一方で、他の局外パラメータ (nuisance parameter) の不確実性は解析において自動的に考慮される。ここで、結論 (信頼区間) は、仮説についての明確な記述というより、信念についての確率論的記述として表わされる。抗糖尿病薬の心血管リスクの評価に関連する多重性を制御する一連のメタアナリシスに対して新規ベイズのサンプルサイズ決定方法が提案されている [249]。

4.4 節で述べたように、事前分布を選ぶことに関連する主観は、否定的な特徴であると一部では考えられる。歴史的に、ベイズ・パラダイムの批評家は、事前分布の選択に対する彼らの批判に焦点を当ててきた。これらの批判には、事前分布を選択する唯一の方法はないという点で、いくつかの妥当性があったが、非情報提供者に対する Jeffreys の貢献は、批判をやや緩和した [250]。より最近では、頑健性と感度分析に関する理論的發展によって、不完全な事前情報に直面したときのベイズアナリシスの基礎がさらに明らかになった。さらに、階層的モデリングの導入により、結果の推論への影響を減らしながら、以前の選択をより高いレベルにプッシュすることが可能になった。

Bender ら [251] は、多重度の問題に対処するための提案をいくつか提供している。彼らは、「体系的なレビューにおける複数の比較の問題に対する単純で完全に満足いく解決法はない。しかし、それは認識を要求する問題である。レビューの著者とユーザーは、多数の統計分析を含む、またはそれに基づいたレビューを提示、解釈、使用する際には、多重性について注意する必要がある」と結論づけている。

この多重性の問題に対する満足いく解決策がないため、累積メタアナリシスは探索ツールとしてのみ使用され、多重性の調整が行われることなく実行した分析の数が単純に報告される。したがって、累積メタアナリシスの多重性へのアプローチの 1 つは、従来の安全性解析の多重性のアプローチと同じようなものであり得る。従来の安全性解析では、仮説検定による名目上の P 値は、通常、限定された数の予め指定された結果に対して調整することなく提供される。

4.8 異質性

主要事項の要約

- 異質性は、一般に、臨床的、方法論的または統計的に分類することができる。臨床的および方法論的異質性は必ずしも統計的異質性をもたらすとは限らない。全般的に低い統計的異質性は、必ずしも特定のサブグループの患者間または研究のサブセット間の治療効果に違いがないことを意味しているわけではない。
- 統計的異質性を検出または推定する主な方法は、フォレストプロットと統計（例：Cochran's Q 統計量、I² 統計量、交互作用の検定）等、グラフィカルなものである。
- 結果の異質性が確認された場合、その異質性に関連する臨床的または方法論的要因があれば、これらの関連性が治療の影響を変える可能性があることを理解し研究することは重要である。
- サブグループ解析およびメタ回帰分析は、異なる種類の結果の探索や、特定の患者または研究の特徴に関する特定の疑問に答える方法である。
- 可能な限り、統計的異質性につながる可能性のある臨床的または方法論的異質性の潜在的な要因はプロトコルに明記され、理想的にはいくつかの特徴は調べられるべきである。
- メタ回帰分析およびサブグループ分析は、注意を払っても誤解を招く結論につながる可能性あり、これらの結論は慎重に記述すべきである。

異質性は、最も広い意味で、研究間および研究結果間の違いを指す。異質性は、一般に、臨床的異質性、方法論的異質性および統計的異質性の3つの方法で分類することができる [191,252]。

臨床的異質性とは、試験間での参加者の選択（試験中の疾病状態、適格性基準、参加者の特性、または地理的差異）、介入（例えば、期間、用量、対照群のタイプ）およびアウトカム（エンドポイントの定義、フォローアップの期間、スケールのカットオフポイント）の違いを指す。方法論的異質性とは、研究デザイン（例えば、ランダム化のメカニズム）および研究実施法（例えば、割当て隠蔽、盲検化、脱落およびフォローアップの欠測の程度および取り扱い、または解析方法）の違いを指す。臨床的異質性と方法論的異質性を構成するものは、計算を伴わない判断に基づいている。

他方、統計的異質性は、個々の研究が互いに数値的に一致しない結果を持つ可能性があり、そのばらつきがサンプリングの変動性のみに基づいて予想されるものよりも大きい、という概念を表す。統計的異質性は、試験間の既知の臨床的および方法論的相違、未知の試験（臨床的または方法論的）の特性によって引き起こされ得るか、またはそれは偶然に起こるものである。図 4.1 は、研究間にわずかな変動しかない（左図）および大きいな変動がある（右図）の2つのケースにおける研究から研究への2変数のアウトカムに対する治療効果の変動性を示す。これらの変動の異なるタイプが統計モデルの選択にどのように影響

するかについては、4.3 節で詳しく説明している。

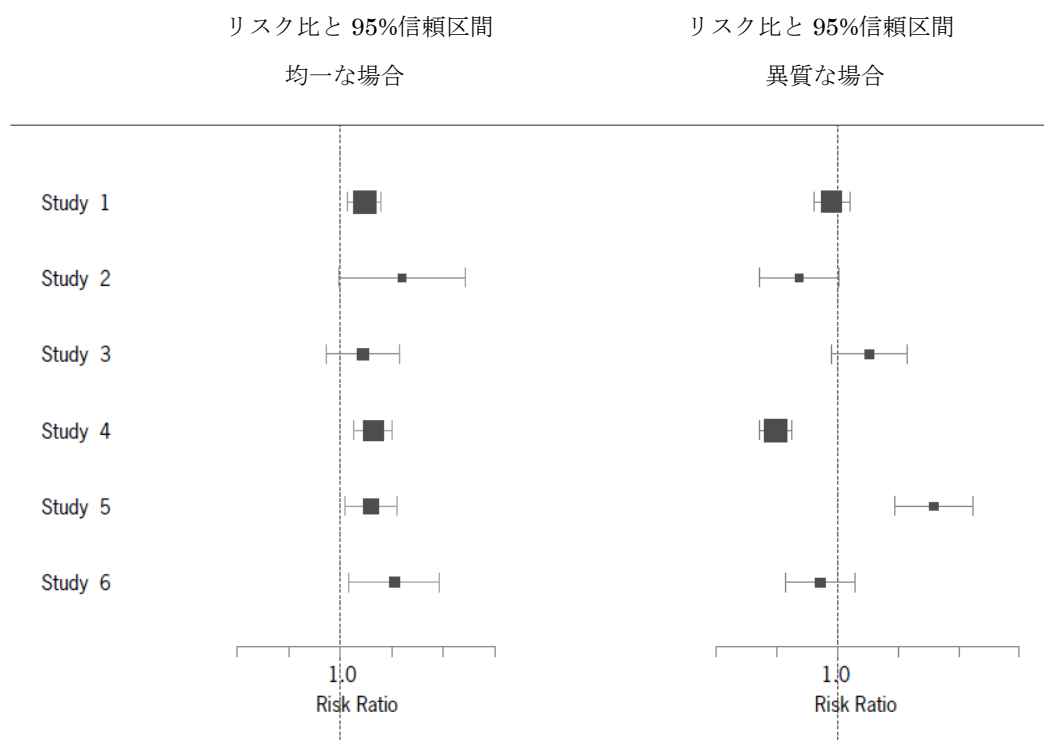


図 4.1 研究間の効果量 (effect size) の違い

出典：Crowe, BJ CIOMS X で作成(2016)

統計的異質性を検出または推定する主な方法は、フォレストプロット (図 4.1 および 4.2 参照) または統計 (例：Cochran's Q 統計量、 I^2 統計量、交互作用の検定) である。Q 統計量と I^2 統計量について詳細は 4.3.4 項を参照。視覚的に図 4.1 の左側の結果は一貫しているが、対照的に、右のグラフからは、何らかの理由 (おそらくバイアスを含んでいる) がある印象を受け、これらの研究は同じ数値を推定していない可能性がある。右の結果は、結果の異質性の理由を理解しようと徹底的な分析を行う動機となる (わずかな研究では難しいであろう)。

統計的異質性がないことが必ずしも臨床的異質性がない、または違う治療での真に治療効果がないということを意味しているわけではないことに留意する。さらに、臨床異質性が必ずしも統計的異質性をもたらすとは限らない。いくつかの状況では、多様な研究における結果の一貫性は、頑健で一般化可能な治療効果を表す可能性がある。

ベースラインの事象率がかなり異なっている研究を組み合わせることは不適切であるとし

ばしば考えられる。しかし、ベースライン（対照）での率の差は、必ずしも研究間の治療効果の大きさに対して大きな差になるとは限らない。さらに、治療効果の異質性が現れる場合、依然として十分に均質であると解釈可能な治療効果をもたらすような指標/スケールを選択することができる。リスクの差は、オッズ比や相対リスクよりも異質である傾向がある[166]。これは、FDA の非劣等試験での業界に対するドラフトガイダンスでも指摘されている[253]。2変数のアウトカムの場合、絶対的な尺度（リスク差のような）からオッズ比または相対リスクなどの相対的な尺度にスケールを変更すると、統計的な異質性が低下し、データが十分に均質になり、その比率スケールで意味のある組み合わせた推定ができようになる。結合された結果は、臨床的解釈可能性を助けるために（例えば、オッズ比(OR) からリスク差(RD)へ）変換することができる。定数 OR モデルは、エフェクトサイズが RD スケールで変化するため、外部データまたは実際のメタアナリシスに含まれるデータからベースライン(コントロール)の事象率を推定するかどうかの判断が必要である。これらの2つのアプローチは、分析、特に分散推定には異なる意味を持つ[167]。

統計的異質性に対処するための方法には、データが正しいことを再度確認すること、全体平均を引用すると誤解を招く可能性がある場合（例：いずれかで null のためエフェクトサイズが劇的に変化する場合など）メタアナリシスを全く行わないこと、異質性を探索すること（例えば、サブグループ解析やメタ回帰分析を用いて）、効果尺度の変更すること（4.1節を参照）、研究を除外することがある[6]。異質性の探索と理解が重要であるため、このセクションの結果は、サブグループ解析とメタ回帰分析にフォーカスする。

4.8.1 サブグループ解析

サブグループ解析には、データをサブグループにわけて、それらの間で比較を行うことが含まれる。研究のサブグループ（異なる地理的位置など）または参加者のサブグループ（男性および女性など）について分析を行うことができる。これらの解析を用いて、異なるサブグループ間の平均エフェクトサイズを比較することができる。サブグループ解析は、特定の参加者グループ、介入の種類や研究の種類、異質性の結果を調査する方法などの特定の論点に答える方法として行われる。

最初に、研究のサブグループを検討する。コレステロールの低下に対する薬剤の影響を評価するためにメタアナリシスを実施すると仮定し、欧州および米国でいくつかの研究が行われたと仮定した。米国と比較して欧州で治療効果が異なると仮定した。2つのサブグループを比較するには、いくつかの方法（固定効果とランダム効果の両方）が使用できる。比較するサブグループが2つ以上ある場合、これらのうちのいくつかは、z-テスト（2つのサブグループ）と分散分析（ANOVA）の研究レベルのバージョンに相当する。例えば、

サブグループ解析は、最初に、メタアナリシス手法を用いて欧州の研究での全体的な平均治療効果（標準誤差を伴う）を計算し、アメリカの研究から別に治療効果を計算するというやり方ができる。次に、平均治療効果が群間で有意に異なるかどうかをテストするために z 検定を用いることができ、差異（2つのサブグループ間の平均治療効果）を計算することができる（標準誤差もあり）。興味のある読者は、研究のサブグループ間の治療効果の差を推定し、これらのサブグループ間の統計的異質性を評価する方法の詳細については、Borenstein ら[4]および Deeks ら[254]の論文でチェックできる。

別々の参加者のタイプに関する十分な詳細はほとんど公表されていないため、研究中の参加者のサブセットのサブグループ解析は文献のシステマティック・レビューでは珍しい。サブグループ特有の結果が提示された場合（例えば、高齢者と若年者の治療効果が別々に報告される場合）、サブグループ特有の治療効果推定は、4.8.3 項で述べた注意事項を念頭に置いてメタアナリシス手法を用いて組み合わせることができる。

対照的に、個別参加者データ（IPD）が利用可能である場合、相当数の解析方法が利用可能になる。方法は、各研究を別々に解析し、標準的なメタアナリシス手法（2段階分析）を用いて効果サイズを組み合わせる方法と1つの手順ですべてのデータを解析する手法（1段階分析）として広く分類できる[4]。後者の方法では、介入効果と参加者レベルまたは研究レベルの特性との間の関連性を探るマルチレベルモデリングなどの複雑な解析が可能である。IPD は、参加者レベルの特性によって定義された参加者のサブグループに対する影響を推定することを可能にする。IPD の1段階分析は各サブグループ内の別々の要約治療効果推定値をだすのにも有用であるが、一般的に、治療のための用語、関心の共変量（サブグループの人数）および相互作用の用語が組み込まれている IPD の1段階分析を行うことがより適切である。しかし、1段階分析を用いても、研究のサブグループ（サブセット）（例えば、フォローアップ期間の長い研究 vs フォローアップ期間の短い研究）から得られた要約推定値を作ることは容易であることに留意されたい。

4.8.2 メタ回帰分析

メタ回帰分析は、解析単位が IPD ではなく研究である回帰分析の一形式である。それは、サブグループ解析の延長であり、連続的な影響およびカテゴリー的な特性の影響を調査することができ、原則複数の要素の影響を同時に調べることができる（ただし、これは多くの場合研究の数が不十分なため不可能）[105]。この手法では、研究の特性（例えば、割り当ての隠蔽、ベースラインリスク、介入のタイミング）と研究効果の推定値（例えばリスク差、ログオッズ比、またはログリスク比）の関係を調べるために研究レベルの要約データを用いる。研究が少ない場合、たとえ患者数が多いとしても、メタ回帰分析は

科学的に有用であるとは考えにくい。メタアナリシスで 10 件未満しか研究がない場合には、一般的にメタ回帰分析は考慮されるべきではない[6]。

メタ回帰分析は、研究レベルの要約統計量を応答変数および説明変数として持つため、回帰の各研究に重み付けし、多くの場合ランダム効果モデルが適切だろうが、適切なモデル（固定効果 vs ランダム効果）を選択することが重要である[255]。多くの方法が考えられており、興味をもった読者は参考文献などの論文として参照している[4、255、256]。

メタ回帰分析モデルは、公表された研究の結果の要約と参加者の特性の要約推定に依存している。これらの要約の結果は、参加者間ではなく、研究間のリスク要因の変動のみを記述している。そのため、研究間では異なっているが、同じ研究のすべての参加者で共有している特性に最も役立つ。

3.9 節で説明したように、フルの IPD が利用可能である場合、相当数の解析方法が利用可能である。

IPD が利用できない場合、研究内の参加者間で大幅に異なる可能性のある多くの特徴は、研究のレベルでのみ要約することができる（例えば、年齢）。これらは、研究レベルで解析することができるが、そのような解析は偏った結果をもたらす可能性がある。この問題は、集約した個人の結果の 1 つであり、集団バイアス、生態学的バイアスまたは生態学的誤謬として広く知られている[139、257、258]。サブグループの違いを検出するための要約レベルのデータの難しさは、生態学的バイアスに限定されない。Lambert らは、シミュレーション研究では、要約レベルの解析ではしばしば相互作用を検出する統計力が低いことを指摘している（効果修正として知られている）[140]。同様に、Schmid ら(2004)は、2 つのデータセットの分析を実施し、IPD 分析によって関心のあるサブグループの差異が検出されたが、サマリーレベルの分析では検出されなかったことを示した[141]。最初の分析では、治療効果は研究全体で均質であったため、メタ回帰分析は相互作用を見つけださなかった。同じ研究からの IPD の解析では、治療効果の重要な修飾要素を発見した。彼らの 2 回目の研究では、少なくとも 10 の研究を用いたメタアナリシスにおける研究レベル要素と異質性のある治療効果または有意義な全般的な治療効果の治療相互作用の検出に有効であるメタ回帰分析が見出された。

4.8.3 サブグループ解析とメタ回帰分析に関する総論

可能であれば、統計的異質性を招く可能性のある臨床的または方法論的多様性の潜在的な原因はプロトコルに明記されるべきである。特性を事前に特定することにより、分析する

要素の数を制限し、研究結果の知識が解析される要素の選択に影響を及ぼさないようにすることにより、擬似的な所見の似たような結果を減らすことができる。2つ以上の特性が調べられる場合、複数の比較を行うために有意水準を調整することが賢明であろう[引用文献[6]、9.6節、5.3節]（多重性の詳細については4.7節を参照）。調べるべき特性は、もし、説得力があるものであれば、臨床的にもっともらしく、他の外部または間接的な証拠によって支持されるべきものである。

特性の事前指定が推奨されているが、研究の定性的な合成を行う際に、レビュー担当者は事前に考えなかった要因を特定する可能性がある。そのような特性は分析することができますが、事前に指定されていない結果を報告することを明確にすべきである。事後（post hoc として知られている）として定義された特性の解析は、データによって導かれる特性の分析である。通常、予備分析の後にのみ示唆されるものである。このような解析は潜在的にデータ浚渫の一形式であり[259、260]、コントロールされないタイプ I エラーをもたらす[259-261]、結果、潜在的なバイアスの対象となる可能性がある。

個別の研究の属性を個々のプロトコルの仕様のみに基づいて特徴づける際には注意が必要である。実際には、現実的な理由から、研究はしばしば元の計画から逸脱している。各研究の実施がプロトコルの仕様を反映していることを確認する努力をするべきである[9、48]。

メタ回帰分析とサブグループ解析は、注意を払っても誤解を招く結論につながる可能性がある。サブグループ解析とメタ回帰分析は、その性質上、観察的である。したがって、他の研究レベル（または参加者レベル）の特性による交絡によるバイアスの可能性を含め、観察調査の限界に苦しんでいる。主な研究グループに与えられた比較可能性が必ずしもサブグループに引き継がれない可能性があるため、観察されたサブグループの差異（たとえ研究内でさえ）が取り違えられることがあり、これらのグループ間の重要な参加者の属性に不均衡が生じる可能性がある[48]。たとえば、性別というサブグループに関心をもつとする。無作為化のために、薬剤と対照薬群の間に性別が均等に配分されるであろうが、もし性別が試験集団の年齢と関連している場合、男性と女性の間年齢が均等に配分されないことがある。年齢が薬剤のリスクをも変える場合、男性と女性の間有害な治療効果の観察される差は、2つのサブグループ間の年齢分布の差によって引き起こされる可能性がある。

異なるサブグループは異なる量の情報を含み、その結果、効果の検出力が異なる可能性が高いため、結果の統計的有意性を単に比較することは大きな誤解を招くことになる。そのような分析を実行する際にはさらに注意が必要である。いくつかの状況では、特に小集団の場合サブグループでは、薬剤群とコントロール群での脱落の違いにより、ベースライン

時の共変量の不均衡が導入され得る。既知および測定された不均衡要素に関するデータが入手可能であっても、研究サブグループにおける測定されていない適切なリスク要因の分布および潜在的な不均衡は未知のままであり、研究結果に予測できない影響を与える[48]。

4.9 感度分析

主要事項の要約

- 感度分析は、メタアナリシスにおける決定と仮定が主要な結果にどのように影響し、それら条件が様々に変動する下、得られる結果がどれほど頑健であるかを調べるアプローチである。
- 感度分析は、メタアナリシスの全段階で行われた決定を反映し、常に統合された推定値の頑健性を評価するために実施されるべきである。
- いくつかの感度分析は研究プロトコルで予め指定することができるが、感度分析が適する懸念事項の多くは、当該研究の個々の特性が吟味されるレビュープロセス中のみ同定される。

感度分析は、メタアナリシスによる決定と仮定が主要な結果にどのように影響し、決定と仮定が様々な条件下でどれほど頑健であるかを調べるアプローチである。例えば、いくつかのメタアナリシス研究では、完全な詳細がないため、適格性が疑わしい場合、感度分析はメタアナリシスを 2 回行うことに関与するかもしれません：最初は全ての研究を含み、2 回目は適格と分かっている研究のみを含む。感度分析では、「結果を得る過程で行われた決定に関して、結果が頑健であるか」という論点がある。様々な専門家は、メタアナリシスの全段階で行われた決定を反映するために感度分析を実施すべきであり[262]、常に組み合わせた推定値の頑健性を評価するために実施すべきであると主張している[263]。

- 感度分析の計画は早い段階から始めるべきである。研究者は、次のような感度分析の重要な決定と仮定を特定する必要がある。
- 除外された研究の特定のタイプを含めること、または含まれている研究の特定のタイプを除外することの効果；
- 組み合わされた推定値が、盲目の使用、割り付けの隠蔽、アウトカムの客観的な確認など、バイアスのリスクに寄与する個々の要因によってどのように影響されるか；
- さまざまな影響尺度（例えば相対的リスクとリスク差）から得られた結果が一致するかどうか；
- 固定効果モデルの結果がランダム効果モデルの結果と一致するか、または固定効果またはランダム効果モデルの異なる形態が一致するかどうか；

- ▶ 欠落データ、ゼロセル、および不完全なデータ報告を処理するための異なるアプローチが結果にどのように影響するか

分布の仮定に対する感度は、調査効果のための異なる分布を仮定し、その後の影響を比較することによって評価することができる。例えば、分析者は、基礎となる研究効果が Student-t 分布から生じ、それによって正規分布から生じるものよりも長いテールを可能にすると仮定することができる。

いくつかの感度分析は研究プロトコルで予め指定することができるが、感度分析に適し多くの問題は、検討中の研究の個々の特異性が特定されているレビュープロセス中にのみ同定される。報告では、そのような事後解析は、そのように明確に識別されるべきである。

固定効果モデルとランダム効果モデルの結果が一致しない場合、その不一致の性質とその理由を理解することが重要である。唯一の違いが精度の点であり、ポイントの推定値が似ている場合は、異質性の原因を探求することが有用である（異質性の 4.8 節を参照）。

このような状況では、精度の違いは、研究内の効果の異質性から生じる可能性が高く、反対方向の変化は「均衡」する。小規模な研究の結果は大規模な研究と系統的に異なる結果を示しているため、ポイントの推定値が一致しない可能性がある（ランダム効果モデルでは、小規模な研究は通常より重く重み付けされる）。もしそうであれば、小規模な研究の特徴、またはそれらの研究に登録された集団は、それらを大規模な研究とは系統的に異なるものなのだろうか？

感度分析により、全体的な結果および結論がレビュープロセス中になされる異なる決定によって影響されないことが示される場合、レビューの結果はより高い確度であると見なすことができる。感度分析により、レビューの所見に大きな影響を与える特定の決定または欠落した情報が特定されている場合は、試験の著者に連絡して IPD を取得するなど、不確実性を解消し追加の情報を得るためにより多くのリソースを割く必要がある。もし、これが達成できない場合は、結果を適切な程度の注釈をつける必要がある。このような知見は、さらなる研究と今後の研究のための提案を生み出す可能性がある。

感度分析は、サブグループ解析と混同されることがある。一般に、感度分析は、方法論的決定による主要結果の頑健性に関係するが、サブグループ解析は特定の研究特性にわたる治療効果の探索に関係する。いくつかの感度分析では、分析を全研究のサブセットに限定することが含まれるが、2つの方法は2つの点で異なる。第一に、感度分析は、解析から除外された研究群における介入の効果を推定しようとするのではないが、サブグループ解

析では、各サブグループについての推定値が生成される。第二に、サブグループ分析では、調べられている要素が治療の効果を变えるかどうかを評価する明確な目的で、サブグループ間で正式な統計比較が行われるのに対し、感度分析では、同じことを評価する異なる方法間で非公式の比較が行われる。しかしながら、原則として、両方のアプローチは、解析に対する一連のアプローチにおいて、結果の頑健性を評価する共通の目標を有する。

4.10 メタアナリシスの報告

主要事項の要約:

- ▶ 従来の文献によるメタアナリシスの報告に関するガイドラインはある。
- ▶ 本章では、規制下の安全性評価における重要な報告項目に焦点をあてた改訂チェックリストを提示する。本チェックリストは Table 4.1 に示しており、主要な4つの項目（導入、方法、結果、解釈）からなる。また、本チェックリストは頻度的、ベイズ解析に特徴的な項目も示している。

メタアナリシスの一貫した取り組みの向上のため、エビデンス統合に関して全ての関係者（規制当局、依頼者、研究者等）がより簡単に理解、解釈するためにメタアナリシス報告の標準化が必要である。標準的な報告はまた、メタアナリシスが他の研究者、関係者によって、再現、保証され得るために、科学的信頼性を向上させ、透明性を示すだろう。フォレストプロットはそのような試験成績の表示における重要な一部分であるため、その観点を主に次項で示す（統計的異質性の推定を背景としたフォレストプロット内容を紹介した 4.8 も参照）。

4.10.1 フォレストプロットの使用

メタアナリシスからのデータの視覚的な表示、特にフォレストプロットは広く使用されており、時間をかけてではあるが着実に極めて標準的なフォーム開発につながっている。フォレストプロットは主要な特性の提示に極めて有用性が高く、たとえ全ての概要を作成しなくとも有用となる可能性がある。Fig.4.2 に例示したフォレストプロットは特徴を示す説明により補足される。

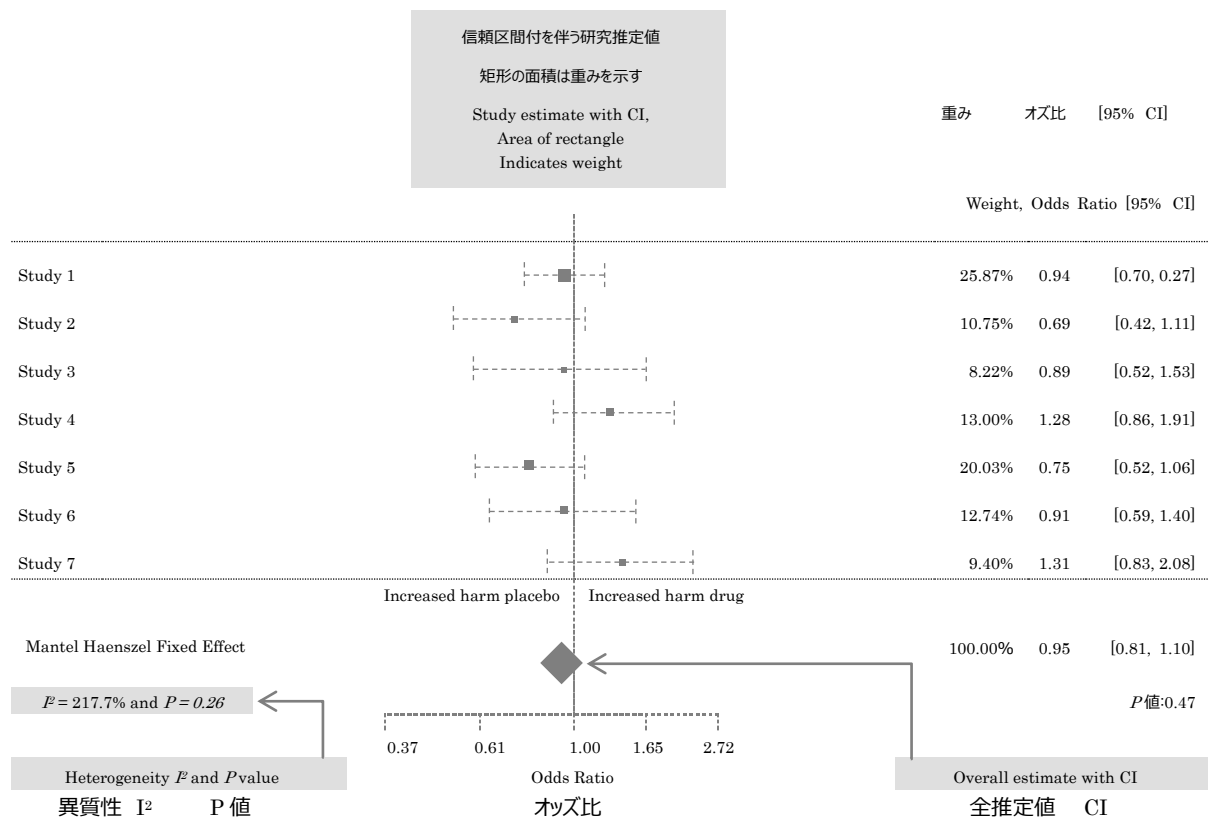


図 4.2 フォレストプロットの一例

出典：Crowe, BJ が CIOMS X (2016) のためにフォレストプロットの例を作図した。

横の線は個々の試験における効果推定における CI (信頼区間) を示している。推定値上にある■ (block) は情報量 (通常、重要度が用いられる) に比例しているか、その範囲がどこに該当するのかわを示している。縦の線は null value を示しており (表現に違いはない)、◆ (diamond) は総合的な概要を示し、試験間の平均的な効果の最適な推定値が中心に示される。

横軸は、効果推定値が OR、RR やハザード比のような相関指標である場合、ログスケールを示すことが多い。このような場合、CI は推定値が対称性を示す。さもないと、それらは非対称となるだろう。フォレストプロットの詳細な歴史や機能は、Lewis と Clarke によって示されている [264]。

4.10.2 報告のためのチェックリスト

従来の文献によるメタアナリシスの報告に関するガイドラインは入手可能である。メタアナリシスの報告に関するガイドラインは、*Meta-analysis of Observational Studies in Epidemiology (MOOSE)*によって2000年に発行されている[265]。2009年には**Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)**ガイドラインが発行されている[18, 266]。これらのガイドラインはそれぞれ、観察研究、RCTのメタアナリシス報告のための推奨事項をまとめたチェックリストを提示している。当該チェックリストは背景、研究方針、選択・除外基準、方法、結果、ディスカッション、結論に関する推奨事項も含んでいる。

さらに、PRISMガイドラインでは、特定したRCT、対象としたRCT、対象外としたRCTの試験数に関する情報及び対象外とした理由を提供するフロー図を提案している（フロー図の例 Fig 4.3 の下部参照）。PRISMガイドラインが発表されてから、複数のガイドラインが追加、発表されている。これらのガイドラインはプロトコル[267, 268]、メタアナリシスネットワーク[20]、個々の参加データ[19]のためのPRISMがあり、また、特筆すべきは、PRISMの具体的な弊害（PRISM specifically harm）[269]も含む点である。PRISM-harmチェックリストは2009年のPRISMチェックリストに4つの項目を追加している。Hammadらは、メタアナリシスの弊害を評価するためのPRISMガイドラインの経験則によるアドヒアランス評価を提示している。また、メタアナリシスの弊害に関して記述をすべきメタアナリシスの特徴を追加した表も示している[270]。

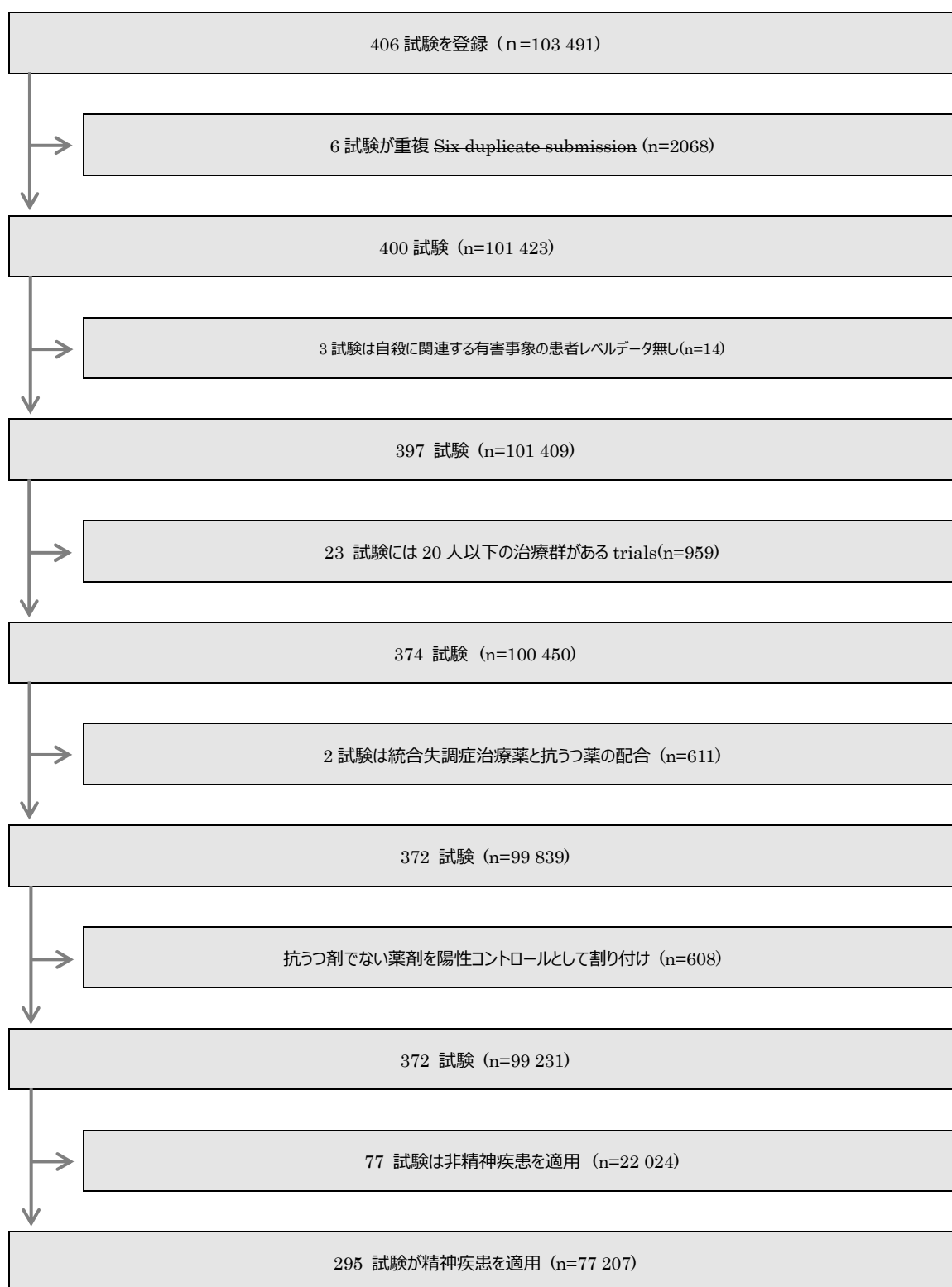


図 4.3 試験および参加者除外のフロー図

出典: Stone et al. Risk of suicidality in clinical trials of antidepressants in adults: analysis of proprietary data submitted to US Food and Drug Administration. *BMJ*, 2009, Fig.1 [271].

文献レビューをベースに、我々は規制下の安全性評価に関して重要な報告要素に焦点をあ

て、メタアナリシス報告のチェックリストを修正した (表 4.1 参照)。本チェックリストは主に 4 つの項目 (導入、方法、結果、解釈) から構成される。各々の項目は、その項目に関するメタアナリシス報告の主要な要素及び記載を含み、各重要な要素に関する簡単な記載から成る。導入の項はメタアナリシスの目的に沿った、規制背景を説明する項目、メタアナリシスの弊害の対応に関する解釈を含む。方法の項ではエビデンス創出の実際の対応に関する項目に焦点を当てている。結果の項では提示すべき特定のデータ概要に関して、である。最後に、解釈に関する項では、メタアナリシスの結果をどのように適切に評価、デジジョンメイキングに使用するかが議論される。

さらに、我々はベイズメタアナリシスという特定の観点も含めた[208]。システマティック・レビュー及びメタアナリシスおよびその過程を考慮した構想 (リサーチクエスチョンをどのように組み立てるか、どのような試験を選択するかといった点) は、解析検討が頻度論、ベイズかのいずれかに関わらず、完全にメタアナリシスの複数部分で相関する[6、5-8 章]。しかし、ベイズメタアナリシスに特徴的な重要要素があり、それは Table 4.1 の初めのカラムにアスタリスクで強調している。

ベイズ法による検討は、医薬品の安全性評価でメタアナリシスを行う点において多くの利点を有する一方、規制下の評価においては、一般的なメタアナリシスの報告に関する透明性維持に通常要求される以上の内容で報告手順を標準化することが非常に重要であるいくつかの理由がある。まず、従来の頻度分析の実施、報告に関する多くのガイドラインや多くの研究結果が存在するが、近年、ベイズ解析に関する合意は統計家のなかで統一されているだけであり、その多くは統計専門組織以外には知られていない。2 点目として、ベイズ法は、よく知られている頻度分析と比較し、複雑な傾向にある。ベイズ法によるメタアナリシスの報告には複数の方法があるため、ベイズ解析を実施する際に報告すべき項目リストの作成が必要となる。3 点目として、チェックリストはベイズ解析に通じていない査読者、読み手の理解を深めるためのガイダンスとなりえる。最後に、ベイズと従来の解析では用語が異なる傾向があるため、標準化されたレポートは、同様の用語を使用することを促進し、ベイズメタアナリシスが報告されたときにはっきりと区別可能となる。

ベイズ解析に重要な要素について述べているガイドラインが少なからずある。Lang と Secic はベイズ報告に関して、以下に示す基本的な要素を提示している。

1. 実施前に処理効果の観点と特定の仮説の蓋然性を報告することとそれらをどのように決定したか特定しておくこと
2. 実施後の蓋然性と対応間隔を報告すること
3. 実施後の蓋然性の解釈

同様の提案は、データ解析実施にベイズ法を用いた原稿著者に対する *Annals of Internal Medicine's* の指示書でも言及されている。The Bayesian Standards in Science (BaSiS) グ

ループ [273]は文献のベイズ解析の報告における標準的な枠組みを作成した。また、Sungらは文献のベイズ解析を報告する際に専門家が最重要と考える7つの項目からなる方法を作成した[274]。Spiegelhalterらによって提唱されたベイズチェックリストはベイズ解析で報告すべき事項に関する強固な基礎を提示した[275]。これらの著者たちは、報告の際にベイズチェックリストを用いた多くの事例を提示している。

表 4.1 メタアナリシス報告のためのチェックリスト（バイズメタアナリシスを含む）

項目番号	メタアナリシス報告の重要な要素	頻度分析フレームワークで考慮すべき事項	バイズフレームワークで考慮すべきこと
導入			
1	導入と背景	該当するのであればメタアナリシス実施の理由と規制背景；関係する母集団の観点と試験への介入性；評価すべき安全性の懸念のエビデンスの叙述に関して知られている事項の簡単な概要	
2	メタアナリシスの目的	リサーチクエスションとメタアナリシスの目的	
方法			
3	想定されるメタアナリシスとは？	データ収集や解析の前にプロトコルの有無を確認する（有であればどこにアクセスすればよいか）。	データ収集や解析の前にプロトコルの有無を確認する（有であればどこにアクセスすればよいか）。データ収集、解析の事前に構成が計画されているかどうか？
4	アウトカム	アウトカム（例えば MedDRA SMQ の範囲やバージョン）及び有害事象の定義が試験ごとに違いがないか定義する。対象となる試験における有害事象の評価期間（例えば、投与期間 + 30 日）を記述する。被験者、処置（投与量）、アウトカム、フォローアップ期間；及びアウトカムの評価手順（処置方法の盲検化等）アウトカムを評価するようにデザインされた試験（すなわち試験中の有害事象が事前に評価項目として定義されているかどうか）と試験終了後にアウトカムが評価されている試験を判別する。有害事象が具体的に収集されるか、それとも自発報告のかどうか。紙報告の有害事象が因果関係ありと判断しただけのものかどうかを突き詰める。判定手順と臨床専門家委員会の憲章（チャーター）を記載する。	
5	適格性基準と試験の選択	適格性のための基準として使用される試験の特徴（例えば、PICOS たフォローアップ期間）を、試験選択の過程を提示しつつ特定する。構成する試験の重複（例えば 1 つ以上の文献で報告されたもの）を避けるため実施した方法を記載する。注目すべき有害事象が全く述べられていないとの理由から除外された試験を明示する。オリジナル試験での母集団の定義のため使用された可能性のある充実した対応を記述する（例えば、心疾患における高い個々のリスクを選択する）。デザインのニュアンスが均一か確認するため、各試験のプロトコルが試験間にアクセスできるか、レビュー可能かどうか特定する。募集した患者の選択の特徴に関して、各試験でプロトコルが実施されたことを確認するために方策がとられたか記述する。	
6	調査方針と情報源	繰り返される使用限界を含めて、少なくとも 1 つのデータベースにおける全電子調査方針を示す。本調査及び最後に実際した時点での全情報源（例えば、対象とした期間とデータベース、追加試験の特定のために著者との接点）を示す。	
7	データ抽出プロセス	文献からのデータ抽出の方法（例えば、調査フォーム、独立性、重複性）、調査者から情報入手・確認プロセスを記載する。	
8	データ項目とリスク測定	調査したデータ（PICOS 等）の全変数、想定、簡略化、変換（例えばリスク測定間での変換）を提示、定義する。	
9	個々の試験および試験間のバイアスのリスク	個々の試験におけるバイアスリスク（試験レベル、アウトカムレベルを含む）の評価を記述する。例えば、構成する試験で実施されたフォローアップ、報告された試験での減量範囲、もしそうであればデータ創出でどのように使用されたか、使用されていないのか、である。	

項目番号	メタアナリシス報告の重要な要素	頻度分析フレームワークで考慮すべき事項	ベイズフレームワークで考慮すべきこと
10	効果測定	選択した効果測定（例えば、RD、OR や RR）、さらに複数の測定方法の転換を実施したかどうかを定義する。	
11	分析母集団	例えば、intention-to-treat、as-treated、per-protocol である。異なる試験では異なる分析母集団となるであろう。メタアナリシス母集団と薬剤毎の患者数のため患者背景データを提示する。表 4.2 に例示する。	
12	エビデンス創出のための統計モデルの特定	モデルレベルの構成；固定効果モデルからランダム効果モデルか、選択のエビデンス；特定の重み付けスキームの選択	事前選択（階層構造の場合はそれよりも事前に）；選択のための分配とエビデンス；感度分析を目的とした代替優先は明確に特定されるべきである。
13	サブグループの解析	（メタアナリシスのための）サブグループ解析を事前に特定しておくか検討する；サブグループにおける処置群間のベースライン共変量の不均衡の可能性を記述する。	
14	ゼロイベント	関連するのであれば、特にゼロイベントがどのように取り扱われているか論じる。	
15	モデルチェックと感度分析	モデルフィットの確認、感度分析を実施するための方法を検討する（例えば、不足データの取り扱いやまばらなデータの取り扱い）	モデル確認と感度分析はベイズ解析に特異なものではなく、ベイズ結果がモデル推定による可能性は高い。
16	コンピュータ計算とソフトウェア	コンピュータで推論構築する方法；使用されるソフトウェアと商業的に入手できない場合どのように保証されているか	コンピュータで後に推論構築する方法；MCMC が使用されるのであれば、どのように収束が確認されるか；その他の関連コマンドとオプション
結果			
17	試験の選択	特定に試験を除外した理由とともに、一次解析の選択のため、選択した試験数を提示し、特定の試験を除外した理由とともにどのように最終試験を選択したか、示すフローチャートを図示する。；有害事象が報告されていないとの理由だけで、どの試験が除外されたかは特に注目する。その文献の著者から追加情報を入手する取り組みを記述し、著者と連絡が取れた場合、返答があったか、もしそうであれば、著者と連絡した結果、どの試験が追加で含まれたか、を記載する。	
18	試験の特徴	各試験においてどんなデータが収集されているのか（試験規模、PICOS、フォローアップ期間等）試験の特徴を示す。	
19	個々の試験および試験間のバイアスのリスク	その試験での、また試験間のバイアスのリスク評価結果を示す。 試験から検出された事項の違いの結果生じるうるプロトコル、試験特性間の違いを（定性的に）検討する。これらは必要に応じ、メタ回帰の中で使用される可能性のある特性である。（例えば、と同じく選択・除外基準の違い、またリスクの高い/低い患者のリクルートにつながる可能性のあるデザインニュアンス、フォローアップ期間、ロケーション（北米/北米以外）、治療環境（入院患者/外来患者）、対象薬群の特質（例えば、実薬比較群の有無）、施設数、試験のサンプルサイズ、スクリーニングでのベースラインの違い、プラセボ反応、処置抵抗性患者、ベースラインの時点で有害事象発現リスクの高い患者、有害事象の既往歴をもつ患者のような副母集団の除外	
20	個々の試験の結果	各介入グループのデータの概要を示し、可能であればフォレストプロットで各試験における処理効果の推定値と信頼・確信区間を指し示す。	

項目番号	メタアナリシス報告の重要な要素	頻度分析フレームワークで考慮すべき事項	ベイズフレームワークで考慮すべきこと
21	結果の創出	整合確保と同様に処理効果、CIの全体像の結果を示す。	全体の処理効果、確信区間、試験特有の後部配置（posterior distribution）の概要（数値、視覚的）及びモデルパラメータ、興味量の全体像を示す。
22	追加解析	実施したのであれば追加解析（例えば、感度、サブグループ、メタ回帰）の結果を提供する。	
解釈			
23	エビデンスの概要	残存する不確定要素と同様に、エビデンスの強さを含むメタアナリシスから検出された主な事項を要約する。可能性のある原因の評価を記載する。	
24	感度分析	結果への感度分析の影響と分析で検出された事項の解釈を含む。	
25	解析の限界性	解析における全ての強み、弱みに関する公平な評価を含む。	

Source: Created by CIOMS X (2016).

Table 4.1 に示す項目の多くは、PRISM ステートメントといった、すでに構築された多くのメタアナリシス報告ガイドラインと重複し、補完するものと考えべきである。

表 4.2 医薬品、医薬品クラス及び治療割り付け毎の患者数

医薬品	主要(一次) Primary	陽性対照 Active control	プラセボ Placebo
選択的セロトニン再吸収阻害剤 (SSRI)			
Citalopram	1928	733	1371
Escitalopram	2567	563	2604
Fluoxetine	9070	2418	7645
Fluvoxamine	2187	0	1828
Paroxetine	8728	1223	7005
Sertraline	5821	1129	5589
Duloxetine	6361	0	4172
Venlafaxine	5693	129	4054
他の現代の抗うつ剤 Other modern antidepressants:			
Bupropion	6018	0	3887
Mirtazapine	1268	0	726
Nefazodone	3319	0	2173
三環系抗うつ剤			
Amitriptyline	0	625	627
Clomipramine	0	632	617
Desipramine	0	315	298
Dosulepin	0	106	95
Imipramine	0	2345	2304
他の抗うつ剤			
Mianserin	0	28	28
Trazodone	0	121	125
全医薬品	52 960	10 367	35 904

出典: Stone et al. Risk of suicidality in clinical trials of antidepressants in adults: analysis of proprietary data submitted to US Food and Drug Administration. BMJ, 2009, Fig.1 [271].

第5章 結果の解釈

5.1 序論

本章の目的は、メタアナリシスの調査結果を評価する思考過程について、最初の直観的解釈から始まり、方法論の妥当性、他の根拠との整合性、また次の規制措置につながりうる調査結果の影響を綿密に評価し、記述することである。本章では、結果の解釈が製品のベネフィットとリスクに関して得られる他の情報と関連付けられること、適切な専門家がレビューに関与することを重要視する。5.2節では、メタアナリシスの臨床上の重要性をメタアナリシスの結果に関わる利害関係者（患者、規制当局、承認取得者）に知らせる必要があるかの緊急性をどのように決定するかを議論する。5.3節は、メタアナリシスの技術的な妥当性を評価する枠組みを示す。5.4節は、妥当な仮定の下で実施されたメタアナリシスの結果の信頼性の評価の仕方や一貫性のある結論を出せるかについて解説する。5.5節は、メタアナリシスの結果をどのように総合的な入手情報に組み込むかについて解説する。また、レビュー中の品目のベネフィット・リスクバランスに関する総合的な意見をまとめるにあたり、異なる情報源からの結果をどのように解釈するかも調査する。この点は、メタアナリシスが、医薬品の添付文書や追加調査遂行の必要性、あるいは追加のリスク最小化策の必要性にどのような影響をもたらすか、について検討する5.6節でも更に議論する。最後に、5.7節と5.8節は、メタアナリシスの結果をどのように報告し伝えるべきか、議論する。5.8節は、コミュニケーションに含むべき情報に関する助言、情報が適切な読者に伝わることの重要性を強調している。また、メタアナリシスの初回レビュー後の暫定的なコミュニケーション、その後より詳細なレビューを提供し、メタアナリシスの結果として製品使用の変更情報を提供するコミュニケーションが必要だろうと力説している。

主要事項の要約：

- メタアナリシスは、医薬品の承認前および承認後のいずれにも実施される。
- 得られた結果の総合的な重要性は、詳細を検討する前に分析されるべきである。
- 適切な専門家がレビューに参加すべきである。
- 結果は製品のベネフィットとリスクに関する他の利用可能な情報と併せて、解釈されるべきである。

メタアナリシスの読者には、それが医薬品承認のいずれの時点で実施されたかに関わらず、臨床医、統計学者、疫学者、あるいは企業、アカデミア、規制当局に属する専門家を含む。メタアナリシスが公表された場合、読者は理論上誰でもなり得るが、本章の目的に合わせ、読者は特に明記しない限り少なくともスポンサーか規制当局を含むと想定する。メタアナ

リスのレビュー担当者が患者への責任を持ち（例えば、承認取得者、規制当局、処方医）、患者の安全性を守るために必要な更なる措置の結果を分析したいとすると、レビュー担当者は、より詳細なレビューに着手する前に、以下の直接的な論点を考慮すべきである。

- 当該メタアナリシスの潜在的な臨床的関連は何か？（5.2節参照）
- 当該メタアナリシスは論文として査読を受け、雑誌に発表済みであるか？（査読されていないものは、嚴重な調査が必要である。逆に、インパクトの高い雑誌に発表済みであることから自動的に精査済みとし、説得力があると解釈すべきではない。実際、公表されたメタアナリシスに基づく調査結果は、それらが他の根拠の数々によって立証されるまでは「仮の」結果として扱うべきである。）

一旦更なる措置の緊急性が決定されたら、メタアナリシスの詳細なレビューにおいて、下記のような問題に取り組むべきである。

- メタアナリシスの技術的な妥当性とは何か？ 5.3節参照。
- 結果はどのくらい強固なものか？ 第3章、第4章及び5.4節参照。
- 証拠全体の力とは何か、また調査結果に対する現存の補強証拠があるか？ 5.5節参照。

メタアナリシスのレビューと分析は、様々な専門性を持つ専門家によって行なわれるべきである。調査結果の臨床的関連を評価するためには臨床、安全性、臨床薬理学の専門家に相談すべきであり、メタアナリシスの技術的な妥当性や方法論に関する問題点を評価するためには統計学や疫学の専門家に相談すべきである。

これらの考察に基づき、意思決定プロセスが開始されるべきであり（5.6節参照）、報告や結果の伝達方法に関する議論を必要とする（5.7節及び5.8節参照）。

5.2 メタアナリシスの潜在的な臨床上の重要性

主要事項の要約：

- 医薬品の承認取得者や規制当局による結果の伝達の緊急性は、想定する公衆衛生上の反響の程度によって影響を受ける。

メタアナリシスの結果公表時、スポンサーあるいは規制当局による何かしらの緊急措置の必要性については、患者や医療関係者に時期を逃さず潜在的なリスクについて伝えるため、両者によって評価される必要がある。レビュー担当者は、メタアナリシスの結果が未知の

リスクや、既知のリスクの頻度あるいは特異性、重症度の増加を見出しているかどうか確認する。該当する場合は、潜在的な臨床上的重要性と更なる分析計画を検討する際に、意思決定プロセスやコミュニケーション戦略と一緒に考察されるべきである。

現時点では、メタアナリシスの方法論的な質はまだ分析されておらず、結果の潜在的な影響のみを考慮することができる。

検討要素は以下を含むべきである。

- 事象の性質と重大性：それは妥当か、回避可能か、治療可能か、可逆的か。注釈：妥当性の欠如は、しばしば単に我々が発現の原因機構をまだ理解していないだけであつたりするので、生物学的妥当性は若干のガイダンスにとらえるべきである。一例として、ジエチルスチルベストロール子宮内曝露と女性の子孫におけるまれな膣がんのリスクの上昇について考える。これは当時説明できなかった事象である。
 - 薬剤への曝露人口の規模、絶対リスク（安全性に関わる事象に潜在的に影響された患者の絶対数）、そして背景となる発生率（もし効果の大きさが割合で報告される場合）
 - 認可された代替治療の利用可能性及び適切性
 - 製品の安全性データ全体における新しい証拠の濃淡と、実質的にベネフィット・リスクに影響を及ぼす可能性
 - エビデンスが子ども、高齢者、妊婦等の特定集団に影響するか
- これらの臨床的な考察や、利用可能なメタアナリシス結果の患者や医師への予想される影響に基づき、緊急性や適切な一連の更なる措置を決定すべきである。

5.3 メタアナリシスの技術的妥当性

主要事項の要約：

- メタアナリシスが既に特定された研究プロトコルに従うかを確認し、メタアナリシスの方法論的なアプローチを綿密に評価する。もしメタアナリシスが査読のある雑誌に公表されていない場合はより徹底的な方法論のレビューが必要である。しかし、公表自体が質を保証するものではない。
- 研究の特徴とメタアナリシスに含まれる全ての研究のエンドポイントの定義をレビューし、メタアナリシスの結果に影響を及ぼす可能性のある潜在的なバイアスの原因を評価する。
- 個別の研究からのデータ抽出の適切性と正確性について吟味する。
- 研究結果における多重性のレベル及び不均一性を考慮する。メタアナリシスの結果を解釈する際、医学的な妥当性や臨床的関連性、他の解析との整合性について評価する。

この節では、実施された分析の妥当性が考慮される。3章と4章は、分析の計画と実行のための適切な方法について詳細に議論し、それぞれ3.1節と4.10節でメタアナリシスの計画と報告の仕方について指示を出している。これらのリストは分析における潜在的な欠落要素や脆弱性を評価するのに参考情報として用いられるだろう。当局に提出されるメタアナリシスの場合は、追加情報を要求する機会があるであろう。本書における既出の資料に基づき、本節では、その結果や解釈が実質的な影響を与える可能性のあるメタアナリシスにおける主要な不備を特定すること、いずれの時点でメタアナリシスが信頼できると見なせるかを確認することに重点を置く。

報告されたデータの正確性については、留意すべきである。可能であれば、複写の正確性と整合性を確保するために、個別の研究データを原データと比較すべきである。

5.3.1 メタアナリシスのための計画の評価

第一段階としては、メタアナリシスのプロトコルが準備されたかどうか、その準備はメタアナリシスのデータが集められ、分析が始まる前にされたのか後にされたのか、そして現状分析はそのプロトコルと一致しているか（解析計画書を含むプロトコルの内容についての3.1節を参照）を判断すべきである。これに関連して、化合物が開発中でも、プログラム安全分析プラン（PSAP; PSAPに関するより詳細な情報についてはCrowe et al. [60, 277] and Xia & Jiang [278]参照）解析計画書として取り扱えることに留意すべきである。本章の後半の意図は、プロトコルや解析計画を参照する際PSAPを含むことである。

分析が行なわれる前に十分に計画されたメタアナリシスは、常により高い信頼性を持ち、様々な種類のバイアスが避けられ、あるいは低減されている。理想的には、計画は可能であれば個々の研究の結果が判明する前に行なわれるべきである。もし最終的なメタアナリシスが元々のプロトコルに従わない場合、プロトコルからの逸脱について、逸脱の理由も含めて記載すべきである。逸脱は比較的まれであるべきで、十分に説明される必要があり、分析結果に基づいて決定されるべきではない。

第二段階としては、メタアナリシスのプロトコルにおける方法論的アプローチを厳格に評価すべきである。多くの原則については、本書の前章、特に3章、4章で記載されている。一般的に、プロトコルは、行なわれる分析の潜在的な多重性やデータ選択、特定の結果に有用な方法を担保すべきである。理想的に、プロトコルは少数の潜在的な評価すべきリサーチクエスチョンに重点を置くべきである。もしあまりに多くのリサーチクエスチョンが順位付けや評価付けをせずに調査されると、計画されたメタアナリシスさえ評価が難しくなるであろう。少なくとも計画を伴う場合、分析の選択肢候補の中から可能な選択の度合

いを定量化することが可能であり、これは無計画な解析では不可能である。

メタアナリシスを評価する細に非常に重要なトピックは、含まれるべき研究の選択である（3.5節参照）。その選択が正しいかどうかを評価するために、読者は様々な状況における多くの例が記載されている3.5.1項を考慮することが推奨される。ここで我々は、意思決定の流れを強固にする鍵となるいくつかの点を強調する。一般に、試験薬の場合には、薬剤の開発者は利用可能な全てのデータ、あるいは対照臨床試験のようないくつかの関連するサブセットにアクセスし、入手することができる。これは理想的な状況であるが、分析を行なう際に文献検索を用いなければならないことや、公表されたデータに限定されることがある。検索方法が包括的であるかを評価することは重要である。例えば、その試みは公表されていない研究からのデータを識別し、集めるために行なわれたのか。その選別された研究は提案された評価に適しているか。もしその効果が長期間服用したときのみ現れると予想される場合、短期間の研究は含めるべきか。多様な効能・効果の研究の組み合わせや、似たような作用機序を持つ複数の異なる薬剤の組み合わせに関しても同様の論点がされるであろう。これらの全てが主観的な評価である。3.2節と3.5節で記載されているとおり、メタアナリシスを適用する際には、多くの起こりうる状況があるので、レビュー担当者は、研究の選択が事前に指定されているどうか、その選択が適切かどうかの両方について検討しなければならない。

計画は評価すべき結果を明確に特定すべきである（3.3、3.4節参照）。臨床試験では、安全性のエンドポイントはしばしば主要アウトカムの測定とは一致しないため、それらは例えば制限のない（非特異的な）患者への質問を通じて、有害事象としてのみ収集されるであろう。しかし、メタアナリシスに利用可能なデータが正しく、完全であるかを評価することは重要である。興味のある情報を患者から引き出せたか、そしてそれは研究のデータベースに記録されているか（我々は必ずしも全ての起こりうる事象を聞き出すべきと主張する必要はない点に留意する。例えば、精神科医は研究のプロトコルが特に要求しない限り、患者に体重について質問しないであろう。しかし、特殊なイベントを聞き出されたか否かは、情報の完全性を評価するのに役立つ）。集団間で確認の差異はあるか。有害事象に関する情報は、研究間で一貫した方法で記録されているか。もしデータが公表物から得られている場合、そしてその研究がイベントを全く報告していない場合、それはイベントが全く起こらなかったためなのか、イベントが収集されたものの報告されなかったためなのか、あるいはイベントが全く聞き出されなかったためなのか。結果は主観的要素を含んだものか。もしその場合、判定あるいは何か他の方法で一貫性を保証しようとしていたか（完全確認の必要性は、より考慮を要するが、全死因死亡率のようないくつかのエンドポイントについては、判定はおそらく不要と言える）。個々の参加者のデータあるいは要約データはメタアナリシスに使われたか。

標準的な、事前に定義された検索方法の一例は、抗うつ薬を投与された患者における自殺傾向（現在は自殺年慮や自殺行動として称される）のリスクを評価するために行なわれたメタアナリシスのセットである。付録1にこれらのメタアナリシスの要約を示す。元のメタアナリシスでは、該当する試験のスポンサーによって、データはアメリカ合衆国FDAに提供され、潜在的な症例を特定するために、標準的な、事前に定義された検索方法が各試験内で使用された。各症例は盲検された独立した読者によって判定された。この過程によって、試験の原データのみから得られる場合と比べて、一貫性や信頼性のあるデータを提供することができる。

最も綿密なメタアナリシスは、サブグループ解析に対する前もって特定された計画や一次解析の感度を評価する分析を含む（4.9節参照）。適切なサブグループ解析と交互作用の分析は奨励されるべきであるが、過剰なサブグループ解析や全集団よりもむしろサブグループの研究結果に集中することは、多重性への懸念を引き起こす。サブグループが計画の中で特定されているという事実は、多くのサブグループ解析が偽陽性の可能性を増加させる（多重性のため）、あるいは偽陰性の可能性を増加させる（特にもし多重性に対して形式的な調整が用いられた場合）といった問題を排除できない。いくつかの問題がある。第一に、サブグループがしばしば本質的に検出力不足であることを考えると、多重性の調整はサブグループ内の治療効果を隠す可能性がある。第二に、サブグループ間の治療効果における真の差もまた、多重性の調整のために見つけられない可能性がある。それでもやはり解析計画で特定のサブグループが特に注意が必要と判断された場合には、より高い信頼性がそのサブグループの結果に与えられるべきである（例:抗うつ薬の例における若年成人）。

4.2節では、イベントが0の研究の問題について議論した。解析計画は、そのような研究が解析に含まれる研究として補足し、適切な解析方法の検討について言及すべきである（イベント0の研究は、抗うつ薬やロシグリタゾン（付録4）の解析ではしばしばあり、論文では詳細に報告されていないが、4.2節で提案されたような追加の解析が行なわれた）。もし特定の研究の結果が、メタアナリシスのきっかけになった場合、その解析計画は結果全体に対するその研究の寄与を評価するためのアプローチを特定すべきである。

全ての利用可能なデータを含む徹底したメタアナリシスでさえ、まだ検出力不足である可能性がある。よって、どの効果の大きさであれば臨床的に意義があり、仮定したメカニズムと一致するか、演繹的に考えることが重要である（4.5節参照）。加えて、仮にメタアナリシスが技術的に確実であっても、間違った結論を出すことを避けるために、1つの試験の結果は他の試験で再現するべきであった[279]。多重性のため、P値と同様に表面上の信頼は誇張されている可能性があるが、信頼区間を調査することで効果の大きさの見込みの範囲について有益な情報が得られるだろう。

5.3.2 正しくない解答を得る可能性（バイアスの可能性、対象集団への調査結果の適応性を含む）

前項では、メタアナリシスの質自体も考慮し、計画段階に関する様々なポイントを取り上げた。これに関連して、質は、当該メタアナリシスから有効な結論を出すための基盤を与える。利用可能な研究の適切性やそれらのバイアスの可能性が問題となり、技術的によく実施されているといえるメタアナリシスは限られている。個々の研究における潜在的なバイアス源に関する厳密な議論は、3.8節に記載されている。研究が文献に基づいて特定される場合、公表されていないデータを見逃し、（4.6.1項で議論されたように）公表バイアスのリスクにつながる可能性がある。特定の研究を検索し損なう理由は他にもあり、例えば、英語以外の研究を含め損なう等がある。エンドポイントの定義の完全性や一貫性は、特に元のデータベースにおける個々の患者データへのアクセスができず、メタアナリシスが要約された結果のみに基づく場合に、確認することが難しい。

潜在的なバイアスを検討することに加え、メタアナリシスの結果をどのようにうまく興味のある集団に適応できるかについても、特に以下の通り評価すべきである。

- 各研究の集団は、メタアナリシスの目的に適合しているか？年齢層、併存疾患、併用薬は、「実際の」集団と比べてかけ離れていないだろうか。
- いつその研究は実施されたか。その間に、研究に影響を与えるような医療行為に、何か変化はあったか？

3.8.4項で議論したとおり、ファンネルプロット[132, 133]は様々な種類のバイアスを検出するために用いられる。公表された研究に基づく解析では、公表バイアスは、学術誌では方向性あるいは結果の統計学的有意性に基づいて研究が公表されやすいこと、著者が調査結果の本質に基づいて研究を提出しないこと[216]が組み合わされて、公表されていないデータを含み損なうことに由来する。肯定的な結果が得られた研究は、しばしば続報で引用されやすく、もし検索方法が包括的なものでなければ、研究の選択に影響を与えているであろう[217]。研究の結果は、研究が多くの雑誌で報告される可能性にも影響を与えるであろう。そのため、メタアナリシスに二重にデータが含まれていないことを必ず確認すべきである[6], [10.1節], [218]。バイアスは、公表された試験内で報告する分析（例：サブグループ）や結果（例：主要アウトカムや二次アウトカムを元の臨床試験プロトコルで特定されたものから変更すること）を選択すること[65, 66]、あるいは計画の変更によってもたらされることもある。これらは、研究プロトコルの中で計画されたが、結果に基づいて公表時に適切と判断されなかった解析や結果、あるいは事後解析として定義され、結果的に公

表された無計画な解析や結果であるかもしれない。よって、公表されたデータを完全には信用せず、もし研究プロトコルが入手できれば、公表された情報と元々計画されていた解析と結果を確認することが重要である。

最終的に、どの範囲までの研究をメタアナリシスに含めるべきかについては常にジレンマがある。以前に述べたように（3.1節参照）、1つの解析は最小限の選択決定に対応すべきである。例：できるだけ原データからデータを用い、メタアナリシスの見解を課すのは最小限とする。そうすることでようやくこれらの見解がどのような影響を及ぼすかあきらかになる。しかし、別の視点では、その目的は、たとえそれが全ての試験のサブセットだけだとしても、できるだけたくさんの無作為の試験でメタアナリシスを実施することであるべきである。これは一次解析を結果の信頼性により厳しいフィルターをかけた研究一式に制限することで事前に特定されるであろう。もし、メタアナリシスの結果から得られる結論に潜在的限界があるならば、感度分析を行なうべきである。

5.3.3 多重性

安全性の問題の評価は、しばしば多重分析や多重検定と関連している。予備の安全性分析の中で、多重性は偽陽性結果のインフレ率や第1種の過誤の制御の必要性と関連している。安全性上の懸念を評価するメタアナリシスにおいて、研究中の安全性に関わる変数を、確認する戦略の一部に含む臨床試験は減多に含まれない。主な理由はそのような一次研究が減多に実施されないからである。加えて、それらはしばしば薬剤のあらかじめ用意される開発計画の一部とはならない。このような状況では、結果は根本的に探索的な性質を持つ。

メタアナリシスの結果を解釈する際、P値の解釈に関連して2つの異なるシナリオが出てくるであろう。まず、実施されたメタアナリシスが明確に薬剤の開発プログラムの一部として計画されている場合、多重検定の影響が事前に議論されており、多重性の調整あるいは検定エンドポイントの階層は前もって特定されている。これは、すべての可能性のある多重性を除去したことを意味するわけではなく、あるいは多重性の調整によって有意性がないと関連づけられた全てのものが真実ではないという意味でもない。しかし、多重性を減らす論理アプローチは慎重に検討され、表面上のP値はこのような状況において解釈される必要があることは予想できる。第二のシナリオは、開発プログラムの一部として計画されておらず、スポンサーや外部関係者によってプログラムの外で実施されたメタアナリシスに関連するものである。この場合、多重性の問題は一般的に演繹的とは考えられず、多重性に関する調整は行なわれていない。そのような解析はデータ駆動型である可能性が高く、他の多重性の問題につながる。1つの安全性への懸念に関する、よく考えられた仮説に基づき、妥当な作用機序によって示唆されるメタアナリシスは非常に説得力がある。そ

れでも、データ収集やデータ駆動型の手法の一部として集められた同じデータ（例えば、特定された安全性上の懸念を提起した公表論文に反応したもの）は、解釈するのが難しい。これは、定められた安全性モニタリングの一部としてメタアナリシスに反対しているわけではなく、統計学的有意性を解釈する上での難しさを強調している。

メタアナリシスで観測された有害事象を評価する際、P値の議論は限界がある。相当な違いを見出した際、観察されたP値の大きさに関係なく、アウトカムの重篤性や重症性に依存して懸念を提起するものであるためである。このように、統計学的な有意性がないことは必ずしも安全性の懸念を排除するわけではない。有効性は単に決定的な根拠が利用可能な時のみ立証されると考えられるが、安全性データの解釈は、有効性の評価と比較して、起こりうる危害に対して措置を講ずるために必要な根拠がより少ない状況で実施される必要がある。統計学的な観点からは、事後解析や多重性に関連する問題を含む可能性はあるが、安全性上の懸念が存在する中、手元にある多数のデータ全体を考慮することで、頻度論的に有意差検定に関する制限があるものの、データは規制措置には十分と考えられる。

反対に、統計学的な有意性は、必ずしも安全性上の懸念の存在を確定するものではない。メタアナリシスの結果は、観察された有効性の推測される重大性、信頼区間とP値（信頼区間とP値は多重性の影響を受ける）を考慮して評価される必要がある。P値の完全に二分された解釈の制限を考慮すべきである。多重検定に対する多重性の調整がない中で、偽陽性の調査結果の確率が増大する可能性を、安全性上の懸念を見落とす重要性と比較検討する必要がある。

メタアナリシスで多重性のレベルを評価する際、エンドポイントの数、サブグループの数、結果の一貫性について検討すべきである。実例として、抗うつ薬を投与した成人における自殺傾向の評価の例を取り上げる（付録1 参照）。図5.1に示した通り、メタアナリシスの全ての結果は自殺傾向を減らす有益な効果に関する控えめな根拠を示した。しかし、若年層患者（25歳未満）のサブグループでは、リスクの大幅な増加がみられ、交互作用に関する統計学的検定を用いることで、この集団は残りの年齢層の集団と一致しないことが示唆された（交互作用に関するP値=0.004）。このメタアナリシスでは、25歳未満の集団はあらかじめ非常に興味のあるサブグループとして同定された。同じ方法論を用いて評価された子供における以前のデータにおいても、同様のリスク上昇の証拠が見られた。よって、現存する証拠と潜在的な年齢の影響に関する以前の仮説は、若年成人における特定のリスクに信頼性を与える。異なる状況での似たサブグループの結果は、偶然によるものと片付けられた可能性がある。

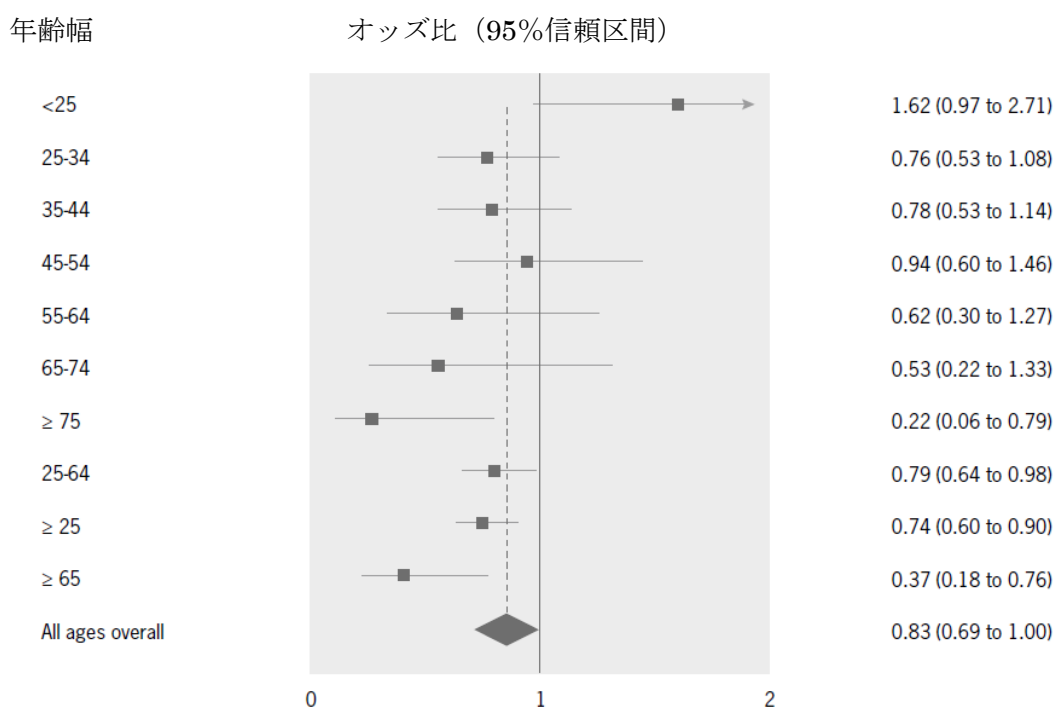


図5.1 精神疾患を持つ成人の年齢相別の自殺傾向（念慮/以上）
プラセボと実薬のオッズ比

出典：Stone et al. Risk of suicidality in clinical trials of antidepressants in adults: analysis of proprietary data submitted to US Food and Drug Administration. *BMJ*, 2009, Figure 2 [271].

5.3.4 治療効果の異質性

4.3.4項と4.8節は、異質性の原因と評価について議論している。含まれた研究間の治療効果に関する異質性の統計学的な証拠は、結果の解釈に必ずしも影響しない。異質性の臨床的関連を評価することは重要である。質的な違い（例：研究間で、効果の方向が異なる）でない、量的な違い（例：研究間で効果の大きさは異なるが、方向は同様である）は、全体の結果の解釈におそらく影響しない。しかし、質的な違いは更なる調査を必要とする。特定の研究デザイン、または関心のある人口要因（より重症な集団vsあまり重症でない集団）がある状況では、不均一性の包括的な検査に頼るよりその要因に関する特定の相互作用試験を行なうほうが適切である可能性がある。これは、不均一性の総合的な有意性の試験がなくても、特殊要素のレベルをまたいで治療効果に関する変動性がある可能性があるためである。

前項（図5.1）で抗うつ薬を服用している患者における自殺傾向の結果と評価は議論され、

フォレストプロットで例示された。もしメタアナリシスが成人と小児両方を含んだ全ての研究を用いて実施されたら、小児の研究（有害な影響を示す）と成人の研究（大部分の成人で全体的に有益な効果を示す）に明らかな不均一性が見られたはずである。

ある程度、不均一性は分析の規模の選択も反映する。大部分のメタアナリシスは、根本的なリスクに比例する治療効果を表す有効性の相対的測定（例、RRあるいはOR）に基づいている。これは、広範囲の基礎的リスクを持つデータを含めることを可能にするが、もし基礎的リスクが異なる場合、相対的な規模で評価された治療の違いは全体の規模でも変わらないわけではないことに留意すべきである。いくつかの例では、基礎的リスクとリスク差（RD）が研究間で相対的に一致している時、全体的なリスクの測定方法（RD）を用いる方がより適切であろう。この差異は重要であろう。というのも、例えば虚血性イベント（心筋梗塞や虚血性脳卒中）のリスクの減少を起こすアスピリンは（同じ作用機序によって）出血性脳卒中のリスクを増大させる。（未治療の）虚血性イベントの基礎的リスクが高い集団では、25%の相対的減少が全体的なリスクの大幅な減少につながった。（未治療の）虚血性イベントの基礎的リスクが低い集団では、25%の相対的減少が全体的なリスクのわずかな減少につながった。しかし、もし（未治療の）虚血性イベントの基礎的リスクが2つの集団間で変わらなかった場合、いくつかの出血に関するリスク要素が虚血性イベントに関するリスク要素と異なるため、アスピリンによるリスクの全体的な増加は集団間で比較的一定である可能性がある。従って、虚血性リスクの全体的な減少は、出血性リスクの全体的な増加と比較して、虚血性リスクの低い集団よりも高い集団で非常に異なる得失評価を示す可能性がある。そしてそれ自体は、これらの2つの患者集団におけるベネフィット・リスクのバランスの評価に影響を及ぼす。

5.4 結果の頑健性

主要事項の要約：

- 結果の信頼性と一貫性を評価する際に考慮すべき点について、解析の中で仮説を生成する場合の限界を含めて議論する。
- 疾患の状態、治療背景、投与レジメン、研究の規模および実施期間など、デザイン上の要因が結果にインパクトを与える可能性がある。
- 評価対象のリスクの尺度となる臨床アウトカムの妥当性と関連性を評価する必要がある。
- 解析判断の性質と多重性、およびエビデンスの強度（*P*値、感度分析の結果）も結果の信頼性にインパクトを与える。

本項では、結果の信頼性と一貫性をどのように評価するかについて、解析の中で生成される仮説の限界を考慮し、議論する。主な懸念事項は、3種類に分類される。a) 対象となる研究のデザインに関する懸念事項、b) エンドポイントの臨床的妥当性に関する懸念事項、c) 解析自体に関する懸念事項、である。4.8項および4.9項で述べたとおり、それぞれの懸念事項は感度分析またはサブグループ解析によってある程度確認することができる。感度分析またはサブグループ解析それぞれについて、解析計画にて事前に規定された解析の方が、事前に規定されていなかったものよりも信頼性が高い。しかし、事前に規定された解析であっても、予定されたサブグループ解析の件数が多すぎる場合は多重性の問題を避けることはできないという点に留意すべきである。ただし、計画された解析判断が全て正しいとは限らないので、解析判断についてはレビュー担当者が独自に評価しなければならない。一般的に、当該手法が妥当であると考えられる場合（レビュー担当者が同じ手法を選ぶかどうかに関わらず）、その一次解析と研究全体の結果により大きな信頼を置くべきであり、感度分析またはサブグループ解析によって相反するエビデンスが得られた場合でも、この基本的な姿勢を変える前にこれらの結果が頑健である必要がある。あるいは、レビュー担当者が特定の解析判断について明らかに不適切であると考えられる場合は、感度分析またはサブグループ解析の方により大きな信頼を置いてよい。例えば、作用機序が共通しているという仮定により「特定のケミカルクラスに属する医薬品を全て含める」という最初の決定が妥当であると考えられる場合は、各々の医薬品に関するサブグループ解析の重みは低下する。異なるレビュー担当者はレビューに対する異なる見解を持つため、複数のレビュー担当者の意見を聞くことは有用である。

5.4.1 研究のデザインに関する懸念事項

5.3.2項で述べたとおり、メタアナリシスを評価する場合、研究デザイン上の要因は重要な懸念事項である。注目すべき特定のデザイン上の要因には以下のようなものを含む（ただし、これらに限定されない）。

- 投与レジメン（ある医薬品クラスに基づく解析をする場合は医薬品）、治療背景、併存症、適応症の違い。
- 研究期間（研究間、または同一研究における投与群間）の違い。特に臨床知識に基づき考えた場合、有害事象が投与開始直後から発現する（即時型アレルギー反応など）のか、それとも長期投与後あるいは追跡調査期間中に発現する（癌など）のかという違い。
- 研究対象集団の違い。
- 対照群または比較対象群が常に同一であるか。プラセボ群または実薬対照群が設定されているか。また治療背景は何であるか。

- ▶ アウトカム：計画された、または計画外の研究にアウトカムの尺度が設定されているかどうか。研究間で定義が一致しているかどうか。必要に応じてエンドポイントに関する判定が行われたかどうか。投与群間に確認法の違いの可能性があったかどうか。
- ▶ 盲検化：アウトカムが主観的なものである場合は非常に重要であるが、死亡率などの強固なエンドポイントの場合はそれほど重要ではない。
- ▶ データの完全性、追跡調査までの中止率／脱落率（自然減少）。
- ▶ 分析単位の問題：解析が行われたレベルは無作為化が行われたレベルと一致しているか。解析において、いわゆる「デザイン効果」の補正が行われたか。

以上の各要因を、抗うつ薬を投与した小児患者の自殺傾向リスクの評価の一部として考えてみる [9]（詳細は付録1を参照）。9種の医薬品に関するデータが含まれていた。これらの医薬品は数種の異なるクラスに属していたが、一般的に自殺傾向に作用すると仮定される機序は特定の薬理学的プロセスではなく、全般的な抗うつ作用である。このメタアナリシスが行われたきっかけはパロキセチンに関する報告であったため、いくつかの懸念事項についてこの医薬品のデータを単独で検討すべきであるが、その他の点では「全ての医薬品が同一の作用を発揮する」という初期の仮説を生成するのが妥当であると考えられる。サブグループ解析を行った結果、2種の医薬品に関しては事象が報告されなかったが、残りの医薬品は全てリスクを上昇させる傾向を持ち、このうち2種の相対リスクはパロキセチンより高かった。パロキセチンを除いたデータの解析結果によっても、全体的な結果が確認された。パロキセチンが属する選択的セロトニン再取り込み阻害薬（SSRI）を投与したサブグループについても結果が得られた。

この事例の場合、4分の3の研究は大うつ病患者を対象としたものであった。残りの研究はその他の精神疾患で、これらのうちいくつかは適応薬として承認されていたのはパロキセチンのみであった。しかし、自殺傾向リスクが上昇すると予想され、また一定の抑うつ症状が発現すると仮定されることから、これらの他の適応症も含める、という事前の判断は妥当であったと考えられる。結果から、全ての研究を対象とした場合も、大うつ病に関するものだけを対象とした場合も、推定される治療効果は同等であることが示された。対照群の全ての症例はプラセボであった。試験期間に関する情報がなく、この点について説明する必要はあるが、通常うつ病を対象としたプラセボ対照試験の実施期間は倫理的懸念から短く設定され、含まれた試験の実施期間に大きな差はないと考えられる。長期的な比較試験が行われないため、リスクの経時的変化を評価することは不可能とされている。

対象となる研究の規模に基づく特定の懸念事項については、3.5項でいくつか議論した。発表済みの小規模試験は大規模試験に比べると、効果量が誇張される可能性があるが、出版バイアスによる可能性がある。変形性関節症に対する様々な治療に関するメタアナリシス

について示した図5.2では、メタアナリシス13件のうち4件でこのような小規模研究効果のエビデンスが認められた ($P < 0.05$) [280]。図5.2から、小規模研究 (1群あたり100例未満) のメタアナリシスと大規模研究のメタアナリシスの間には平均治療効果に差があることが、推定値とCIによって示されている。サブグループ解析の結果、こうした効果を示すエビデンスを示す場合は、効果全体がバイアスによるものであるのかどうか、また残りのデータから効果 (当初の推定値より小さいものであっても) を示す十分なエビデンスを得ることができるのかどうかについて、レビュー担当者は検討しなければならない。小規模研究が効果量の増大に関連することをデータが示唆していたとしても、これは難しい判断である。このバイアスがなぜ生じたかについて決定的な説明ができない場合は、このリスクの定量化に関して疑わしいとしても、患者の安全という利益を守るために、リスクは依然として存在する、という立場をとる必要がある。小規模研究バイアスを補正するための方法がいくつか提案されており、この方法による感度分析はバイアスの潜在的な大きさを評価するのに役立つ [181, 281]。

それぞれの「介入」は、疼痛関連のアウトカムに関するメタアナリシスである。「効果量の差」は、小規模 (1群あたり100名未満) 試験のメタアナリシスと大規模 (1群あたり100名以上) 試験のメタアナリシスの間の、介入に関する疼痛スコアの差を示している。「相互作用のP値」は小規模研究のメタアナリシスによって得られた疼痛スコアに対する効果量と、同じ介入に関する大規模試験のメタアナリシスによって得られた疼痛スコアに対する効果量との間に差があるかどうかを確認した結果を示している。対象となった13件のメタアナリシスにおいて、大規模研究と小規模研究の間に認められた効果量の差の平均値は -0.21 (95%CI: $-0.34 \sim -0.08$, $P=0.001$) で、これは小規模試験が示す有益な効果の方が平均して高いことを示唆している。

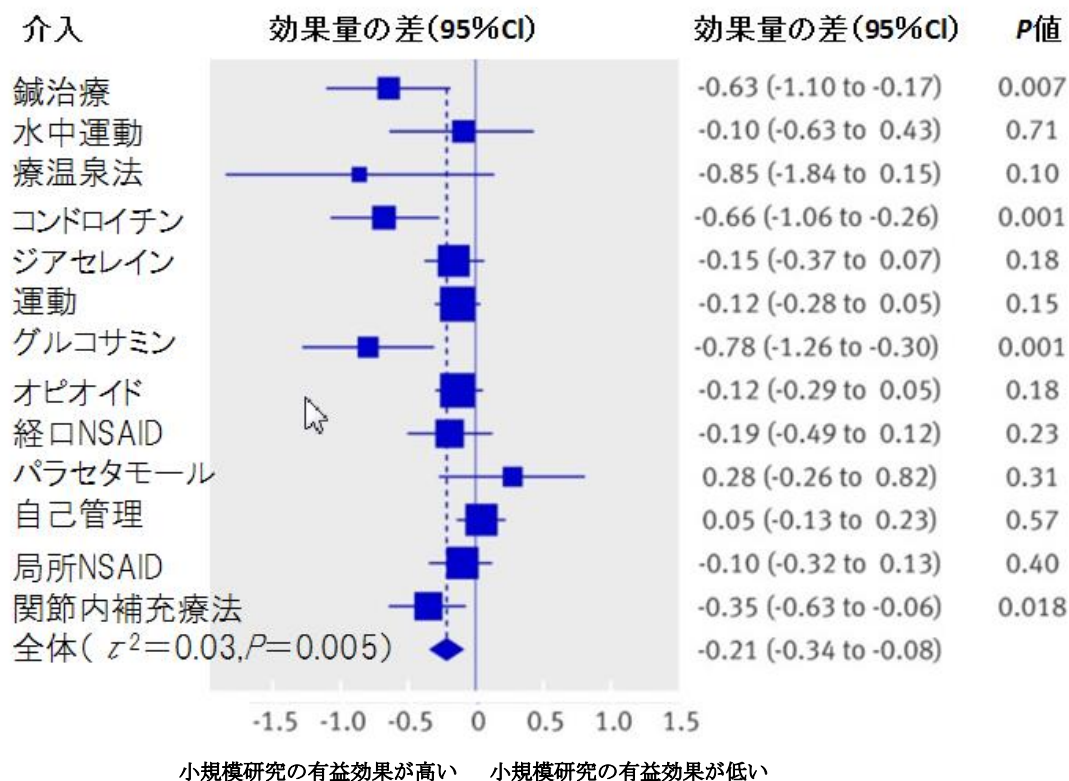


図5.2 変形性関節症のメタアナリシスにおける小規模研究効果

出典 : Nüesch et al. Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study. *BMJ*, 2010, 341, Figure 1 [280] .

5.4.2 エンドポイントの臨床的妥当性に関する懸念事項

ここで臨床的妥当性に関する懸念事項にフォーカスをあてる。この要因はより主観的な性格を持ち、これまでに述べたものよりも、追加解析にて説明することが難しい可能性がある。対象となる臨床試験の実施期間を考えた場合、試験対象の治療とアウトカム間に因果関係があると考えerことは妥当だろうか。例えば短期試験では、成長の遅い腫瘍やアテローム性動脈硬化症の進行に対する因果効果が検出できる可能性は低い、血栓または急性腎不全などの事象に対する因果効果は検出できる。

試験が行われた時期も考慮すべきである。何年も前に行われた試験には最新の治療が反映されていない可能性がある。これには治験薬（化学療法薬の投与レジメンや治療薬の組み合わせは時間とともに大きく変化する可能性がある）や治療背景の経時的な進化が関係している場合がある。可能であれば変化を遂げた特定の要因を直接調査すべきであるが、個別のデータにアクセスできない、または複数の要因が関わっている場合は時間が有用な

代替要因となりうる。時間とともにリスクが上昇または下降することを示すエビデンスについては、全体的なベネフィット・リスク評価の中で重み付けを行うべきである。しかしながら、慎重な解釈が必要である。出版年が有用な代替要因となりうるということは、時間もまた、複数の研究単位の要因と強く関連しているということでもあるため、ばらつきの原因を解明するのはチャレンジングである。

重症度が異なる事象を組み合わせたものをエンドポイントとする場合（小出血と大出血など）、各々の事象の個々の評価を含めて、エンドポイントの臨床的妥当性についても考慮しなければならない。長期的予後が明らかなアウトカムの場合でも、サブグループ間で発現率が大きく異なるアウトカムを評価するのは、やはり複雑だと考えられる。

既述のとおり、自殺傾向の評価などではより複雑になる。自殺既遂または自殺未遂は明確なアウトカムであるが、小児試験のメタアナリシスの場合こうした事象はなく、成人の場合でもこのような事象は少ない。従って、解析の大部分は自殺念慮の事象の解釈に依存する。念慮の増加は実際の自殺企図のリスク上昇と関連するという考え方は論理的に思われるが、このテーマに関する研究領域にはまだ明確な方針がなく、患者の精神状態に関する正確な情報を得ることの難しさにより複雑化している。また、自殺率は年齢と性別によって大きく異なる。特に自殺傾向のリスクが高いことが特定されているグループ（小児および若年成人）の自殺既遂率は最も低い。これは、患者が実際に持つリスクを特定することは非常に困難であるが、念慮は自殺行為に至るために必要な段階であるため一定の自殺既遂リスクは存在する可能性があるとみなす（一つの意見であるが）ことは良識的であると考えられる、という一例である。

5.4.3 解析自体に関する懸念事項

最後に、解析自体に関するいくつかの要因も頑健性に関わっている。初期の結果を論理的にフォローしてゆく中で解析計画からの逸脱が生じることが予想される。こうした逸脱は比較的少なくすべきであり、また明確に説明できる必要があり、個別の研究または初回メタアナリシスの結果によって生じたものであってはならない。結果が出てからアウトカムの定義や重要性を変更すると、バイアスのあるミスリーディングな結論が導き出される可能性がある [282]。しかし、状況によっては予期しない結果または外れ値によって解析を計画する過程では検討されなかった合理的な仮説が生まれる可能性もある。例えば、予め規定された試験の性質または集団によって結果の異質性を説明する試みがすべて失敗した場合を想像してほしい。観察データを綿密に調査すれば、将来的に検討またはモデリングを行うための仮説を生成できる可能性がある。効果量が多い試験は全て一定の時期以前に行われたもので、その時期は投与レジメンの特定の側面（併用薬など）が変更され

た時期と一致することが判明することもある。この仮説はデータを観察した結果生成されたものであるため正式な仮説試験は無効であるが、関連する臨床領域においてこの仮説を試験される可能性がある。

一次解析において、結果の頑健性は P 値にある程度反映される。もし P 値が非常に小さい（0.001未満など）場合は、様々な感度分析またはサブグループ解析を行ってもリスクがいくらか上昇するという結果に大きく影響する可能性は低い。しかし、治療効果の点推定値と分散の両方を検討して、この点を緩和しておく必要がある。 P 値が小さい場合、効果量が小さく分散が極めて少なければ感度分析に対して頑健であるが、感度分析によって根底のバイアスの原因について説明できなければ、明らかに頑健であることによる再確認はミスリーディングとなる可能性がある。同様に、多重性を補正しても結論が変わらないため、多重性に対する懸念もまた小さくなる。累積メタアナリシスにも特有の懸念がある（4.7項参照）。多重性の問題を処理するための手法に広く認められているものがないため、結果が逆転することは珍しくないということをレビュー担当者は少なくとも認識しておくべきである [283]。

サブグループ間に一貫性がないかどうかも調査して、このことが解析の根本的な問題になっていないことを確認する必要がある（解析の規模の選択、および効果の異質性との関係について述べた5.3.4項の議論参照）。4.8.1項および4.8.3項ではサブグループ解析にまつわる問題の多くについて述べた。一般的に、メタアナリシスに含まれる単一のサブグループから得られる結果の信頼性は低いが、この結果が真実でありうる場合もあることは認識しておくべきである。多重性が確認されていなければ、レビュー担当者は調査したサブグループの個数に応じて非公式に P 値を修正してもよい。例外は、過去に得られたいくつかのエビデンスから、特定のサブグループに対して既に推測的な懸念がもたれている場合である。例えば、抗うつ薬を投与した成人を対象とする自殺傾向の調査（付録1、[271] および図5.1）の場合、既に小児に対する効果を示すエビデンスがあるため、解析計画では対象とする主要サブグループは若年成人と規定された。このように、一次解析の仮説は全ての成人にリスクがある、であったが、年齢との相互作用がある、また若年成人の方がリスクは高い、という結果は信頼できるものであり、他のデータとも一致している。

4.3節では固定効果モデルとランダム効果モデルの選択について述べた。多くの場合、ランダム効果解析の方が固定効果解析に比べ導き出される P 値が若干大きく（重要性が低く）CIの幅が広いため、感度分析を行えばそのような差が生まれることが予想される。解析間の差が大きい（効果の点推定値に大幅な差があるなど）場合は、研究間で重み付けが異なると重大な影響が出るということを示す。4.3.1項で述べたとおり、いずれの解析法の信頼性がより高いのかを結論付ける前に、小規模研究の結果を注意深く確認するべきである。

5.5 エビデンスの統合と解釈

主要事項の要約：

- 解釈においては、安全性の懸念事項の臨床上、医学上の妥当性および患者への影響を考慮する必要がある。この要因は症状および生活の質または量への影響も含み、また反応または事象の予測可能性、予防可能性、可逆性によって決まる。
- 対象となる安全性の懸念事項に関するメタアナリシスの重要性は、研究の限界およびその外的妥当性も同様に考慮して決定しなければならない。メタアナリシスの重要性は、主にメタアナリシス自体の品質と、診療現場で当該製剤を投与した集団に結果がどの程度当てはまるかによって決まる。
- メタアナリシスの結果は、得られた他の関連情報（非臨床データ、他の臨床試験および観察試験から得られたデータ、自発報告など別の解析から得られた結果）も含めた全体の中で統合しなければならない。それぞれのエビデンスのソースは、適切性、品質、妥当性について個別に評価し、また妥当性と一般化可能性について慎重に評価した後、安全性の懸念事項に関して得られたデータ全体と概念的に統合しなければならない。有効性に関するデータを裏付けるエビデンスを提供する異なるソースの強度は、必ずしも安全性に関するデータにおけるエビデンスの強度と必ずしも一致しない。

メタアナリシスおよびその他のソースから挙げた安全性の懸念事項の評価には、メタアナリシス自体に含まれるデータの評価以外にも様々な側面がある。いかなる安全性の懸念事項であっても観察された安全性の側面の臨床上および医学上の妥当性を考慮しなければならない。5.5.1項では、5.2節で述べた高レベルの情報に加えて、これらの側面について詳しく述べる。メタアナリシス自体の様々な側面は、そのメタアナリシスの妥当性、頑健性、限界、外的妥当性によって評価する。これらの側面の大部分については5.3項および5.4項で述べた。しかしメタアナリシスによって生成されたエビデンスを、得られた全てのエビデンスの重みと併せて全体の中で重み付けするために、これらの側面が重要になるため、この問題に関するいくつかの高レベルの側面については本項で述べる。5.5.2項にメタアナリシスから得られたエビデンスの重み付けにとって重要な側面の概観を述べる。評価中の副作用について全体的な結論に到達するためには、安全性の懸念事項に関するエビデンスの異なるソースを評価しなくてはならない。また、安全性の懸念事項全体を解釈することができる。5.5.3項に異なるエビデンスのソースがどのように全体的評価に統合されるかについての概観を述べる。こうして安全性の懸念事項に関して利用可能なデータすべてについて、ベネフィットに関して得られたデータを考慮して評価することがベネフィット・リスク評価につながる（5.8.2項を参照）。

5.5.1 安全性懸念事項や患者への影響における臨床上、医学上の妥当性

5.2節で詳述したように、メタアナリシスに用いられた手法を理解する前に、予備的評価を行うが、安全性懸念事項の臨床的重要性は、有害事象が患者の症状、生活の質または量に及ぼす影響に依存する（軽度の消化管症状は消化管大出血事象、永続的な神経障害、突然死よりも重要性が低いなど）。しかし、医薬品による副作用の発現リスクの受容もまた、予想されるベネフィットおよび治療対象疾患の重症度と死亡率に依存し、これらの評価がベネフィット・リスク評価の一部となる（5.8.2項）。

臨床上の妥当性もつ側面として、これ以外に反応または事象の予測可能性、予防可能性、可逆性がある [284]。これらの側面が全てメタアナリシスにて説明されるわけではないこと、またこれらの点をさらに解明するためには他のデータも考慮しなければならない可能性があることを留意すべきである。

事象の予測可能性は、発現リスク、具体的には反応の発現率をどの程度予測できるか、生じうるリスク要素が発現リスクにどのように影響するか、そして個人またはサブグループの特性によって投与の副作用をどの程度修正することができるかに関係している。例えば、特定の有害事象のリスクは特定のサブグループの患者においてのみ上昇する可能性がある。これは自殺傾向の事例に該当すると考えられ、若年患者のみが抗うつ薬によりリスクが上昇することが示されている。可能性のあるリスク要素を知ることにより、標的とするリスク軽減活動を可能にすることができる範囲内において、予測可能性と予防可能性は密接に関連している。この点について以下に述べる。

予測可能性の問題を定義する一つの方法は、安全性の懸念事項を評価する際に、考慮すべき事象の絶対リスクだけでなく相対リスク（それぞれの推定値の精度を含めて）を検討することである。投与群のリスクを比較対象群と比べたときに、相対リスクの推定値は高いが絶対リスクが非常に低い場合は、投与群の中で少数の患者はこの有害事象に影響を受けている、ということを示唆している。有害事象の発現頻度の公衆衛生上の重要性は一般的に絶対値に関連しているが、メタアナリシスの場合は相対値を用いて解析し、結果を報告してもよい。従って、相対リスクと併せて必ず絶対リスクを検討しなければならない。しかし、相対リスクの絶対リスクへの変換には一定の限界（タイムスケールに対する依存など）があることを考慮しなければならない。また絶対リスクはベースラインリスクによって決まるが、これは集団およびサブグループによって異なる。

予測可能性の二つ目の側面は、安全性の懸念事項を評価する場合、特殊な集団のサブグル

ープに対する潜在的な異なる効果を示すエビデンスを考慮する必要があることである。ここで意味する異なる効果とは、有害事象リスクに対する投与の相対的な効果がサブグループによって異なるという意味である。異なる集団を評価するメタアナリシスから関連するデータが得られる可能性もあるが（安全性の懸念事項を評価するメタアナリシスなど）、エビデンスが別の観察データやその他のデータソースから得たものである可能性もある。評価対象のリスクに関係する可能性がある患者特性についても考慮すべきである。例えば、年齢、疾患の重症度、肝または腎機能障害、その他の併存症、遺伝的多型、妊娠、授乳などの特性は治療による発現リスクを上昇または低下させる要因となる可能性がある。

予防可能性とは、リスクを低減できる可能性と関連している。アクセスの制限、例えば、ある副作用に関する内因性の絶対リスクが比較的高い（複数のリスク要素を持つなど）特定の患者集団には投与しないということだけでリスクが低減し、これにより副作用の予防可能性がある程度達成できる場合がある。例えば、ある医薬品がある有害事象のリスクを倍増させることが判明したとする。投与しなくてもその事象のリスクが1%ある患者の場合、そのリスクは2%に上昇する。内因性リスクが10%の患者の場合はそのリスクが20%まで上昇する、つまり絶対上昇値が大幅に大きい。メタアナリシスでは患者集団に関するこのような懸念についても説明することができ、適切な解析を行えば有害事象のリスクが相対的に高い集団と低い集団を判別するための有用な情報を得ることができる。これ以外にも、摂取法または投与経路を変更することによって予防が行える可能性がある。この側面には薬理的または生物学的側面が関係しているため、臨床試験のメタアナリシスではほとんど用いられることがない。こうした薬理的または生物学的側面が試験に組み込まれていない限りは、適切なサブグループ解析を行ってこれらの側面に対応することができる。

副作用の可逆性という側面はメタアナリシスによって説明できる可能性があるが、通常はメタアナリシスとは別の、臨床または前臨床データに基づく薬理的および臨床的側面が関係している。予防可能性の場合と同様に、解析を行うためにはこれらの要因に関連するデータが臨床試験において収集されている必要があり、また適切な追加解析を決定しなければならない。あるいは、前臨床データまたは初期段階での臨床データの複合解析を行うことができる場合もある。この解析については本報告の範囲外である。

考慮すべき重要な要因は、当該製剤が一般集団および特殊集団における使用規模であり、このことはリスクが相対的に高い集団または低い集団に使用することができるかどうかを評価する必要があることを示している。ある医薬品を使用できる集団の規模を知ることによって、潜在的に影響を受ける患者の実際の人数を明確に把握することができる。またこのことは公衆衛生に与える影響を評価する上でも重要である。

有害事象の医学的重要性は、別の治療選択肢があるかどうかによっても決まる。これらの治療選択肢はそれ自体が別の有害事象と関連している可能性があり、また安全性プロファイルの特性が明確になっていないために副作用の発現率に確実性がない場合もある。場合によっては、これらの有害事象はメタアナリシスによって特定されたリスクよりも重症度が高いことがあり、影響を与える患者の数が評価中の実際のリスクよりも多い可能性がある。投与の中止によって考えられる影響は、重大なこともあれば、予後がそれほど悪化しないこともある。緊急的な医学的必要性による投与の中止の方が、評価中の副作用よりも予後が悪い可能性もある。例えば置換治療（インスリン依存性糖尿病の場合のインスリンなど）を中止することは、ほとんどの臨床状況において不可能である。有効な抗癌治療を中止する場合には慎重なベネフィット・リスク評価を行う必要がある。しかし、減量のために用いられている医薬品の投与中止（別の非薬理的介入を行う場合）は、前述した状況下での投与中止ほどの重大な結果にはならないと考えられる。

評価中のリスクが用量と投与経路にも左右される可能性があるため、リスク期間と投与期間の関係についても考慮しなければならない。これらの点もまたメタアナリシスによって説明されており、同様に、効果の時間依存性または用量依存性もサブグループ解析やメタ回帰分析に基づくものとなってきた。しかし、メタアナリシスに未承認の高い用量が含まれる場合は注意が必要である。用量依存性を解析すればシグナルの妥当性は判明するが、承認用量より高用量の場合、解析結果が全面的に臨床診療に適用できるわけではない。時間依存性の側面は、追跡調査期間とリスク期間の長さが異なる可能性があるため（医薬品によって誘発または促進された癌は投与後数年経過してから検出される場合があり、また検出される時期は通常、メタアナリシスの対象となった試験の追跡調査期間には含まれないなど）、メタアナリシスの際には慎重に評価する必要がある。用量依存性の試験に関するもう一つの警告は、柔軟な投与レジメンを用いる試験についてである。3.5.1項で述べたとおり、参加者は有害事象の発現に応じて異なる用量に変更することができるため、医薬品の投与と所定の有害事象との間に認められる関係が正確ではない可能性があるため、注意すべきである [48]。

安全性の懸念事項の重要性は、有害事象に関する既存の知見と新たな知見によってさらに明らかになる可能性がある。規制当局の見地からすると、この重要性は有害事象リスクが製品情報に含まれているかどうか、また既にリスクマネジメント計画または定期的安全性最新報告／定期的ベネフィット・リスク評価報告に述べられているかどうかによって判定される。

これらの懸念事項はベネフィット・リスク評価の対象となる（5.8.2項）。

5.5.2 安全性懸念事項の研究におけるメタアナリシスの重要性の重み付け

メタアナリシスの結果の解釈には、研究の外的妥当性の程度だけでなく解析の限界を考慮しなければならない。5.3項および5.4項では以下の内的妥当性と頑健性について詳述した。

- ▶ 当該メタアナリシスの計画の妥当性。
- ▶ 当該メタアナリシスに含まれる臨床試験から得られた研究データの品質。
- ▶ 試験間および試験内サブグループ間における結果の異質性。
- ▶ 予想される有害事象の時間的経過（既知の場合）に関連する追跡調査期間の長さ。

評価対象の安全性の懸念事項に関するメタアナリシスの重要性と妥当性を批判的に評価するためにさらに考慮すべき点について、以下に述べる。

メタアナリシスの重要性は、比較対象が適切に選択されているかどうかによって決まる可能性がある。解析に含まれる試験には、プラセボもしくは実薬対照が用いられることもある。投与群のリスクがプラセボ群よりも高くなる場合もあるが、実薬対照が用いられ、その対照群も同じ有害事象のリスクが高い場合は、治験薬と実薬対照の間に差は検出されない。従って、プラセボ群と実薬対照群を同時に用いることには問題があるが、異なる比較対象について個別にメタアナリシスを行うことは非常に有益である。このメタアナリシスを行う場合は、多重試験に関する問題を考慮し、また*P*値についての二分された解釈における結論にのみ頼るのではなく、入手できたデータを綿密に評価しなければならない。

メタアナリシスの妥当性は、そのメタアナリシスに含まれる集団の代表性によって決まる。医学診療における集団は臨床試験に組み入れられた集団とは大幅に異なる可能性がある。臨床診療においては、有害事象のリスク、または投与によるリスク上昇に対する感受性は、臨床試験で観察されるものとは異なる可能性がある。試験の選択または除外基準によって患者が選択されてしまう結果に起因している。例えば、メトホルミンは乳酸アシドーシスのリスクを高めるため重度腎機能障害の患者に対しては禁忌である。

従って、メトホルミンと他の医薬品を比較評価する臨床試験の場合は、重度腎障害の患者が除外される可能性が高い。これらの臨床試験のメタアナリシスに乳酸アシドーシスのリスクが存在しないことは、重度腎障害患者にこのリスクがないことを意味するものではない。投与による相対リスクが集団間で異なること、あるいは試験対象集団のベースラインリスクが小さかったためにシグナルを検出できるだけの十分な統計学的検出力（注目事象が十分であること）が得られなかったこと、のいずれかの理由によって、シグナルが見逃

された可能性がある。

5.5.3 安全性の懸念事項を取り巻く状況

安全性の懸念事項を解釈するためには、メタアナリシス自体に関連するデータとは別に安全性の懸念事項の状況を考慮する必要があり、またメタアナリシスによって得られたデータをその他に得られた情報と併せて全体の中で統合しなければならない。

これは、その他に得られた情報とは非臨床データ、単一の介入試験または臨床試験、観察研究から得られたデータ、他のメタアナリシスの結果、自発報告システムから得られたデータである。異なる状況においてソースの強度も異なるため、一般的にあるデータソースが別のデータソースよりも優れているとみなすことはできないことに留意すべきである。エビデンスの情報源は個別に評価し、また妥当性と一般化可能性を慎重に評価した後、そのシグナルに関して入手できたデータ全体と概念的に統合しなければならない。本報告はメタアナリシスに焦点をあてているため、メタアナリシスから得られた結果の評価基準についてのみ詳しく記載する。しかし5.5.2項に述べた考え方は、原則的にはエビデンスのソースの評価に適用できるが、ただし評価対象のデータの性質によってはいくらかの修正が必要である（例えば、観察データの場合は未測定の交絡に関する問題について十分に評価することができるが、非臨床データの場合は研究が適切なガイドラインに従って行われたかどうかを評価することになる）。

このようなエビデンスのソースの評価とは、検討中のシグナル（およびこれを示す利用可能なエビデンス）の薬理学的生物学的妥当性の懸念事項を含めるものであり、観察された効果を説明できる作用機序が存在するか否かを確認することである。しかし、生物学的機序は安全性の懸念事項が観察または確認された後に初めて調査、検討されることが多い。言い換えれば、メタアナリシス（または他のソース）から得られたシグナルに、生物学的機序を発見するための実験が導かれるということであって、その逆ではない。検討すべきその他の事項は、有害事象に関して観察された時間的な関係とその妥当性である。

安全性シグナルに関して得られたエビデンス全体について検討する際も、安全性の問題に関する既存のエビデンスと、シグナルが検出された後に生じたエビデンスを区別しなければならない。例えば、メタアナリシスまたは発表済みの疫学研究の結果が、その後生まれるデータに影響を及ぼす可能性がある（研究対象の効果に関する薬理学的理論に促されて生じた自発報告や仮説など）。特に米国では、初期の結果が訴訟および盛んなメディアパブリシティを生み、その結果自発報告が急増する。

他のデータソースを評価する場合、有効性に関するデータを裏付けるエビデンスのソースの強度は、必ずしも安全性に関するエビデンスの強度と一致しないことを考慮すべきである。エビデンスの強度はデータまたは研究の品質（評価対象の問題に対処する方法の適切性）によって決まる。また、これはデザインそのものではなく、研究の内的および外的妥当性と一般化可能性によっても決まる。品質が高く有効性を裏付けるエビデンスであるとみなされたデータでも、特定の安全性アウトカムに関して重大な欠陥がある場合もある（希少事象の発現率を評価するには症例数が不十分な場合、または特定のリスク要素を持つ集団が研究から除外されている場合など）。一般的に、有効性はピボタルな無作為化対照試験を行うことによって確認されるが、特定の安全性の側面を説明するにはこれらの試験だけでは不十分な場合がある。これらの側面を評価するためにはさらに研究を行う必要がある。例えば、フルオロキノロンと腱障害、あるいはガドリニウム含有造影剤と腎性全身性線維症との関係は、自発症例報告および臨床試験から得られた当初の情報によって評価することはできない。

臨床試験データおよび優れたデザインのメタアナリシスから得られた安全性の問題に関するエビデンスは、一般的には強度が高いと考えられているが、特定の状況下では観察データや別のデータソースの方が妥当性は高いとみなされ、評価全体に極めて重要な役割を果たすことがある。例えば、研究対象の安全性事象が非常に少ない場合や、臨床試験に組み入れることが困難な集団が影響を受ける場合（静脈血栓塞栓症、網膜剥離などの安全性事象、または新生児、妊婦に関連する安全性の問題など）が該当する。これらの場合において、臨床試験データがないことは、この事象が存在しないことを示しているのではなく、この安全性の問題を調査するにはデザイン（無作為化試験）に限界がある（特定の安全性事象を適切に確認するには被験者数が少なすぎるなど）ことを意味している可能性がある。

このような状況では、方法が適切でバイアスおよび交絡に対して頑健である場合に、観察データなど別のデータソースが高強度のエビデンスとなる可能性がある。データおよびエビデンスのソースは全て適切性、品質、妥当性について評価する必要があり、メタアナリシスの結果はデータ全体の中で解釈しなければならない。無作為化試験のメタアナリシスから得られたもの以外のデータでも、以上の留意点に基づいて全体の中で解釈を行えば、より重要性が高まり、より重要な役割を果たす可能性がある。

5.6 医薬品の使用に与える影響

主要事項の要約：

- ▶ メタアナリシスの知見は、医療専門家向け添付文書の改訂、リスク最小化計画、追加試験の要請など、医薬品に関する多くの規制活動に影響を与える可能性がある。

- ▶ メタアナリシスの知見は、新たな安全性情報とみなすことができる。
- ▶ 医薬品スポンサーは、彼らの医薬品の1つに関連する臨床的に重要な知見を有する新しいメタアナリシスを認識した場合、関連する規制当局に通知しなければならない。
- ▶ メタアナリシスの知見は、その医薬品について入手可能な全ての情報と併せて解釈すべきである。
- ▶ メタアナリシスの実施または知見に関連する不確実性の解決には、追加の情報、データまたは追加試験が必要となる場合がある。

メタアナリシスにより臨床的に重要かつ従来は知られていない治療効果についてのエビデンスが得られる場合、これにより製品の使用方法の変更が必要になるであろうし、さらなる情報が必要となる可能性がある。規制当局、医薬品スポンサー、医療提供者、患者、医療制度、保険者など、様々な関係者が決定を下すことが求められる場合がある。本節の焦点は医薬品使用許可の承認において直接の役割を果たす規制当局および医薬品スポンサーとなる。最も重要な注意はメタアナリシスの知見に特有の決定プロセスである。しかしながら、メタアナリシスの知見に関連する一般的な安全性の知見のプロセスにも注意が払われている。

他の情報源から得られた知見と同様に、メタアナリシスから得られた安全性の知見により、最終的に、承認されたラベルへの警告の追加、ラベルにおける承認された用法の変更、医薬品へのアクセス制限、医薬品安全性監視の強化、教育努力および追加試験を含む一連の規制措置につながる可能性がある。また、製品のリスク管理計画に変更が必要な場合もある。規制当局と医薬品スポンサーの両方は、安全上の問題を考慮し、具体的な措置を強制または提案する内部諮問機関を持つことができる。場合によっては、メタアナリシスの背景に基づき、措置に関する勧告を行うための公的諮問グループを招集することができる。

メタアナリシスに関連する規制の背景には他にも意思決定プロセスの要素がある。メタアナリシスの知見は、医薬品ラベルの変更や追加試験の実施を含む、医薬品スポンサーの特定の行動を必要とする新しい安全性情報と考えることができる。これらの措置はスポンサーによって開始されるであろうが、規制当局によって要請されたり、要求されたりすることもある。そのため、スポンサーと規制当局との間の早期かつ完全なコミュニケーションが不可欠である。欧州では、このようなコミュニケーションは、「新たに発生した安全性の問題」の概念を導入したグッド・ファーマコビジランス・プラクティス第VI版（セクションVI.C.2.2.6参照）によって取り上げられており [285]、可能な限り早急に規制当局に通知しなければならない安全性上の問題に分類される。医薬品スポンサーは、医薬品の安全性に影響を与える可能性がある、公開されているメタアナリシスまたは別の規制当局に提出されたメタアナリシスの存在を規制当局に知らせることを期待されている。

一般的に、安全性の知見は薬理学、前臨床試験、主要なランダム化臨床試験、安全性に特化した試験、自発報告、登録データ、観察研究および医薬品クラスに基づく医薬品の情報および知識の集積の一部である。情報と知識は絶え間なく進化している。決定プロセスに必要な情報は、(1) メタアナリシスの設計と実施の妥当性の評価、(2) メタアナリシスの知見の背景の評価である。これらのトピックは前章および本章の前節の主題であり、安全性アウトカムの重大性、アウトカムに関して推定される医薬品の影響、潜在的に危険にさらされている人々の数、医薬品の全体的なベネフィット-リスクプロファイル、その知見がその医薬品の安全性プロファイルに関する潜在的に新しい知見を示す程度を含んでいる。メタアナリシスの妥当性を評価するためには、メタアナリシスに関する臨床専門家や統計的専門家を含め、必要な専門知識を持つチームを利用する必要がある。メタアナリシスの背景を評価するためには、医薬品使用の専門家や医薬品のリスクとベネフィットに関する特定の研究分野の専門家を含む追加の専門家が必要となる可能性がある。

メタアナリシスの妥当性と背景の評価はお互いに決定プロセスに影響を与える。重要な限界を伴うメタアナリシスでさえ、背景の評価を考慮して特定の決定につながる可能性がある。例えば、臨床現場での使用に関する情報が限られている新たに市販された医薬品については、大きな不確実性を伴うメタアナリシスにより、安全性コミュニケーションの保証や医薬品安全性監視の向上につながる可能性がある。

メタアナリシスの限界の特定および対処により、メタアナリシスの知見の不確実性を低減し、より効果的な意思決定を行うことができる。メタアナリシスの不確実性を解決するために費やされるリソースは、メタアナリシスの背景の評価および解決により意志決定が変更される可能性を考慮に入れなければならない。規制当局と医薬品スポンサーは、メタアナリシスの著者からの追加情報を要求するか、または研究報告書に提示された情報に基づいて追加解析を行うことによって、公開されていないであろう追加情報を取り扱うことができ、さらにメタアナリシスの不確実性の一部に対処することができるであろう。プロトコルと解析計画が公表されなかった場合、規制当局と医薬品スポンサーは研究著者にこれらを要求すべきである。また、不確実性や方法論的懸案事項に対処するための追加の要約と解析を研究著者に依頼することもある。

限られた公的に利用可能なデータに基づくメタアナリシスに関して、規制当局は医薬品スポンサーがより徹底したメタアナリシスを行うよう要請することができる。医薬品スポンサーはおそらく患者レベルの試験データへアクセスするだろうし、場合によっては公表されていない試験結果にアクセスする可能性がある（すべての新しい試験が公表されることを期待するであろうが）。患者レベルのデータでは、必要なデータが利用可能であれば、妥

当性と一貫性に関して元の試験で使用されたアウトカムの定義が改良される可能性がある。さらに、異なる曝露パターン、事象の時間パターン、およびサブグループを、一貫性のためまたはリスクの高いグループの特定のために探索することができる。医薬品スポンサーは彼らが実施したり、寄与したり、または認識している試験を含め、追加の試験を特定することができる。

規制当局は、単一の医薬品のメタアナリシスを関連する医薬品のクラスレビューに拡大することを選択することができる。これはクラス全体の安全性のアウトカムを理解するため、または安全性の問題に関する利用可能なデータを増やすために実施することができる。機密保持と妥当性の懸念から、規制当局は徹底的なクラス全体のメタアナリシスを行うことができる唯一の当事者であろうが、Askling らによるメタアナリシスの事例[210]と同様に、場合によっては、規制当局やスポンサー以外の中立的な当事者が関与する可能性がある。

5.7 結果報告における注意点

主要事項の要約

- ▶ 試験選択と患者選択の方法は、報告書の重要な部分である。
- ▶ メタアナリシスの結果を報告するときは、相対的効果量に加えて絶対的効果量を提示する必要がある。

メタアナリシスの報告書に登場すると予想される項目のチェックリストが表 4.1 に提供されている (4.10.2 項に示されている)。

フローチャート (4.10.2 項の図 4.3 など) は、メタアナリシスの最終的な試験と患者集団がどのように選択されたかを示している。

各メタアナリシスの報告書の一部に、メタアナリシス集団と医薬品ごとの患者数の概要を説明すべきである (4.10.2 項の表 4.2 はそのようなプレゼンテーションの一例である)。

本書の前半で、フォレストプロットはメタアナリシスに含まれるすべての試験の中で関係する安全性エンドポイントの効果量を比較するのに適したツールであると説明した。加えて、それは単一およびメタアナリシスの効果測定の結果、試験内および試験間のばらつき、および各試験がメタアナリシスに寄与する重みを上手に要約している。個々の試験内において、エンドポイントやサブグループまたはその他の解析が選択的に報告されている可能性についても考慮すべきであることに注意しなければならない。フォレストプロットやデ

ータの欠測に関する詳細情報やガイダンスについては、4.10.1 項と 4.6 節を参照すること。

大多数のメタアナリシスにおいて、効果量は相対リスクまたは OR として示されている。しかしながら、公衆衛生の観点からのリスクの解釈については、絶対的效果量は絶対リスクに対する治療の影響がその医薬品が使用される様々な患者集団のベースラインリスクに応じてどのように変化するかについてのある尺度とともに提示されるべきである。データの可用性と試験間での効果量の不変性を仮定する意志に応じて、複数のメトリック（OR や RD など）を使用して解析を実行するかまたは 1 つの尺度（たとえば OR）で解析を行い、それから集約尺度を別のメトリックに変換する。Localio らは、OR を他のメトリックに変換する方法を提案している[167]。リスクの認識はそれが提示される方法によって影響され、絶対的基準での提示は、患者と医療専門家の両方によってより正確に解釈される[286]。例えば、心房細動を治療するために使用されるワルファリンは、虚血性脳卒中のリスクを 60%低下させるが、出血性脳卒中のリスクは 200%増加させる。このような提示方法は、ワルファリンの使用が有害であることを示唆している可能性がある。しかしながら、出血性脳卒中の危険性が虚血性脳卒中のリスクよりもはるかに低いと理解されれば、虚血性脳卒中の絶対リスクは 3%低下し、出血性脳卒中の絶対リスクは 1%増加することが理解できる。

5.8 結果のコミュニケーションにおける留意点

主要事項の要約

- メタアナリシスから安全性の知見を伝達するプロセスは、リスクコミュニケーションの他のプロセスと密接に関連している。このトピックに関する指針と報告書を利用可能である（[287]など）。
- リスクコミュニケーションは、すべてのステークホルダーが患者の安全性を最重要視して責任を認識しなければならないプロセスである。知見の解釈の透明性と結果に関連する不確実性も考慮し、安全性の知見を明確かつ正確に報告する必要がある。
- 適時適切な措置の実施を可能にするためには、適切な時間に適切な受け手に情報を提供する必要がある。さまざまな受け手（聴衆）の個別の懸念に応じて、異なるレベルの詳細と説明が必要な場合がある。
- 安全性の問題の解釈プロセスには、異なる評価のステップが含まれることがあり、異なるシナリオの伝達につながり、早期のコミュニケーションを伴う可能性がある。完全なレビューが完了する前の、早期のコミュニケーションは、暫定措置を実施する必要がある状況、または現在医薬品による治療を受けている患者ができるだけ早く服用を中止すべき状況において重要である。早期のコミュニケーションが提供される場合、

情報のその後の解析から生じる重要な情報は、適時伝達されなければならないことに注意することが重要である。

- ▶ メタアナリシスの結果を伝達するときは、相対的効果量に加えて絶対的効果量を提示すべきである。

結果の事実報告とは別に、メタアナリシスの結果を伝達する適切な方法を考慮する必要がある。メタアナリシスから安全性の所見を伝達するプロセスは、他のリスクコミュニケーションプロセスと密接に関連しており、このトピックに関して利用可能な指針と報告書がある（例えばウプサラモニタリングセンターによる医薬品安全監視に関するダイアログ、欧州医薬品庁のグッドファーマコヴィジランスモジュール XV - 安全コミュニケーション [287-289]）。リスクコミュニケーションの全体的な話題に関する徹底的な議論は、この報告書の範囲外である。したがって、この節ではメタアナリシスに関連するリスクコミュニケーションの重要な原則の議論に焦点を当てる。

安全性の知見のコミュニケーションには、規制当局、バイオ医薬品スポンサー、ジャーナリスト、研究者、医師および患者など、さまざまなステークホルダーが関わっている。当面、各々が安全性の知見に関して異なる展望と異なる興味を持っているであろうが、リスクコミュニケーションはすべてのステークホルダーが患者の安全を最重要視して責任を認識しなければならないプロセスでもある。この責任は結果の解釈における透明性と結果に関連する不確実性をも考慮して、安全性の知見を明確かつ正確に報告することを求めている。加えて、情報が最初に判明したときに、情報の完全な報告に影響を与える可能性がある他の当事者との事前の合意など、いかなる制約も宣言する必要がある。そのような状況には、アウトカムデータの実際の所有者でないか、完全な情報の提供が禁じられている状況が含まれる。

安全性の知見の伝達に関係する関連ステークホルダーについては、タイムリーかつ適切な措置の実施を可能にするために、適切な時間に適切な受け手に情報を提供する必要がある。情報伝達が必要なさまざまな受け手（例えば、医療従事者、公衆またはメディアのメンバー）に応じて異なるレベルの詳細および説明が必要であり、透明性は提供される詳細レベルとは独立して確保されるべきであることに注意しなければならない。

5.8.1 解析の異なる段階におけるコミュニケーションの留意点

安全性の問題の解釈プロセスにはさまざまな評価の段階が含まれる場合があり、評価の段階に応じて異なるシナリオの伝達につながる可能性がある。目前の安全性の問題に基づき、かつ透明性の観点から、評価結果のさまざまな伝達段階を計画し、タイムリーな情報が提

供されるようにする必要がある。

完全なレビューが完了する前の早期のコミュニケーションは、暫定措置を実施する必要がある状況、または現在医薬品による治療を受けている患者ができるだけ早く服用を中止すべき状況において重要である。しかしながら、現在の状況がこのカテゴリーに該当するかどうかを判断するにあたっては、明確なレベルの判断が必要であることに注意する。暫定的な結果が提示された場合、情報の暫定的な性質を前もって述べるべきであるが、暫定情報は状況の完全な透明性を維持するために提供されており、利用可能になった場合にはさらなる更新が提供されることに注意すべきである。最終的な解析ははまだ実施されていないにしても、その時点までのデータレビューは安全性の問題を示しており、医師にとっては患者の治療選択において再評価が必要となる可能性がある。(さらに、伝達が必要な安全性上の問題の構成に関しては個人により異なる閾値を有する可能性がある。)

潜在的な混乱を避けるために、関係するリスクを明確に説明し、医師と患者の両方に適切なガイダンスを提供する必要がある。中間結果の不適切で軽率な伝達により問題が生じる可能性があることに注意しなければならない。状況によっては、中間解析により、医療提供者と相談することなく患者が治療を中止することにつながるであろうという、特有の損害のリスクが増加する可能性がある。治療中止に伴うリスクも相当なものになる可能性があるため、患者と医療提供者との対話が不可欠である。患者がある医薬品から同じクラスの他の医薬品に導かれる可能性もあり、結局、完全な調査の後に、同じリスクに関連付けられる可能性がある[290]。重要なことに、中間結果が後続の解析で裏付けられていない場合は、あらゆる中止の結果を回避できたであろうが、それは結果論でしかない。したがって、早期のコミュニケーションが提供される場合、情報のその後の解析から明らかになるさらに重要な情報は、医師と患者の両方にタイムリーに伝達されなければならないことに注意することが重要である。

5.8.2 メタアナリシスの情報を組み込んだベネフィット・リスク評価の留意点

規制当局は、新たな介入のために肯定的なベネフィット・リスクバランスを強調し維持することにますます焦点を当てている。最近の議論は、ベネフィットがリスクを上回るかどうか、不確実性に対処する方法、およびリスクを管理する方法について、より透明な意思決定プロセスを目指している。メタアナリシスがベネフィット・リスク評価に関連するデータ全体からの情報を表していることは注目に値する。メタアナリシスは異なる試験設定で収集されたすべてのデータの選択に基づいて、単一の安全性の問題を評価する場合がある。しかしながら、ベネフィット・リスク評価のプロセスはしばしば単一のメタアナリシスの評価を超え、それによって他のデータソースを考慮し、リスクとベネフィットの両方

を評価する。したがって、メタアナリシスの結果から安全性の知見が明らかになり、裏付けられた場合、影響を受ける医薬品の重要なリスクに関するすべての情報は、医薬品の主なメリットとの関連において提示する必要がある。安全性の知見の頻度、関連する危険要素、副作用発現までの時間および可逆性（および回復までの時間）に関する情報とともに、安全性の知見の重篤性および重症度に関する利用可能な情報を提供する必要がある。

ベネフィット-リスク評価の結果は、ベネフィットおよびリスクに関する不確実性および限界、可能性のあるリスク最小化、およびこれらの問題の重みおよび重要性を明確に示すことによって伝達されるべきである。十分に確認されたアウトカムに関する要約レベルのメタアナリシスは、既定のレベルのリスクを排除するのに十分であり、医薬品の好ましいベネフィット-リスクプロファイルを裏付けることができる。対照的に、医薬品のベネフィット-リスクプロファイルに実質的に影響を及ぼす、これまでに同定されていない、潜在的な安全性のシグナルは、メタアナリシスの不確実性を解決するために大きな努力を必要とする可能性がある。メタアナリシスの結果がベネフィット-リスクバランスの評価に影響を及ぼしたかどうか（影響したならそれはどの程度まで影響したのか）、また、早期のベネフィット-リスク評価と比較してこの評価に関する変更があるかどうかを検討しなければならない。異なる使用条件でのベネフィット-リスクバランスがどのように影響を受けるかということと同様に、メタアナリシスの結果に基づく重要な最新情報を示す必要がある。

5.8.3 メタアナリシスとベネフィット・リスク評価から生じるガイダンスの留意点

結果の評価を提示することとそれを医薬品のベネフィット・リスクプロファイルの観点に置くことに加えて、安全上の懸念にどのように対処するかについての医療専門家および患者への勧告はすべて明確に述べるべきである。関連する製品の使用または使用を継続すべきかどうか、および治療を変更する必要があるかどうかについてステークホルダー（病院、診療所の患者または一般の医師など）に具体的なガイダンスを提供する必要がある。代替治療のリスクおよび関連する医薬品の使用回避又は使用中止に関わるリスクを考慮し、明確に伝達する必要がある。代替療法への安全な切り替え方法についてのガイダンスが必要な場合もある。

該当する場合は、提案された製品情報のいかなる変更に関する情報も提供する必要がある。入手可能であれば、参考文献と詳細な情報源を明記する必要がある。

5.8.4 著者、ジャーナル編集者と規制当局の関係の留意点

ジャーナルに投稿された原稿は、出版前において明らかに機密であるが、著者やジャーナル編集者は、特に医薬品の安全性の知見に関して、データの適切かつバランスの取れた普及に関する責任を認識すべきである。安全性データのタイムリーな評価と適切なコミュニケーションプロセスを保証するために、著者およびジャーナル編集者は、医薬品の安全性の知見が公表される前に規制当局に通知することを考慮する必要がある。研究成果は、機密保持契約の下で公開前に提出される可能性がある。ジャーナルと著者の両方からの報告が同じ研究成果に言及している場合は、そのことを規制当局が確実に理解しているかについて注意を払うべきである。

5.9 結び

関連する安全性データのメタアナリシスの結果は、医薬品による治療中の患者の安全性を守るための適切な措置を開始するために慎重な解釈が必要である。

本章では、解釈プロセスがどれほど複雑で時間がかかるかを示した。しかしながら、構造化アプローチ（5.1 節参照）は、エビデンスの統合、意思決定プロセス、そして最終的には適切な伝達の手順を用いてメタアナリシスの査読者を導く。

メタアナリシスが行われたときの医薬品のライフサイクルの段階および潜在的な臨床的関連性に応じて、さらなる措置の緊急性および解釈の複雑さは大きく変わる。

第6章 リソース

メタアナリシスは大規模プロジェクトとみなされるべきである。非常に早い段階で徹底的な計画を立て、専門知識を運用し、技術的リソースを特定することは、良好なアウトカムを達成するうえで重要である。第6章の目的は、有意義なメタアナリシスを実施するうえで必要なリソースに焦点を当てることである。特定の論点には特定の専門知識が必要となるため柔軟な対応が奨励されるが、チームの意思決定の手引きとして一般的な参考文献を記載した。科学的に計画を立てるだけでなく、チーム専用のレポジトリを用意し臨床試験実施基準（good clinical practice, GCP）の文書化要領に準拠することが推奨される。適切な研究戦略および適切なデータソースがメタアナリシス成功の重要な基準となるため、本章ではそれらの点から入門ガイダンスを行う。メタアナリシス案の範囲には、公開データソースとともに専有データソースについても適格で利用可能な限り、検討することが推奨される。

6.1 コアチーム – メタアナリシス実施の組織的側面

メタアナリシスを成功に導くためには、多様な技術および専門知識が必要である。したがって、産業、規制、または学術分野において、適切な技術を有するコアチームを設立し、メタアナリシスのデザイン、実施および/または評価における全課題の完遂、ならびに最終報告から生じる方策の補助、促進に協力して取り組むべきである。必要な技術は実施するメタアナリシスによって異なるが、少なくとも医薬品安全性、統計、疫学、調査と関連する臨床の現状、一般的な臨床医学、プロジェクト管理、および情報研究（熟練の専門司書からの情報など）における専門知識をメタアナリシスチームが有する、あるいはアクセス可能であることが推奨される。また、専門知識（毒性学、薬理学、バイオマーカー、データ分析プログラミングなど）が必要なときは、外部の人間をプロジェクトに招く柔軟性を有するべきである。

コアチームの第一の課題のひとつに、メタアナリシスの計画がある。最初の計画段階の一部として決定すべき詳細事項には、メタアナリシスの目的または目標に対する合意、組み入れ対象研究の種類、検索対象のデータベースまたはその他の情報源、プロジェクトの対象時間帯および主要管理点、アウトカムの普及戦略、課題および責任の割り当て、および意思疎通および指揮命令系統の明確化などがある。

実践的な注意点として、チームはメタアナリシスに関連する全情報のデータレポジトリを

手配すべきであり、プロジェクトの記録管理と透明性の要件を考慮して、保存する情報と一貫性のあるラベル付けについて検討すべきである。複数のスポンサーからの多数の試験に関与するシナリオ（規制当局からの要請時など）については、要請された変数の明確な定義とコード化方法の明確な定義を有するマスターデータファイル（共通データモデル）を開発することが不可欠である[9]。

分析の際は、使用する統計ソフトウェアをチームで選択すべきである。結果に基づいて決定を下すのではなく、そのソフトウェアが適切な統計方法を有しているか否かで決定すべきである。

6.2 一般的リソース

メタアナリシスの計画を立てる際に役立つリソースは多数存在する。これらのリソースや、メタアナリシスの適切な計画および実施に役立つ他のガイダンス文書にチームメンバーが注意を払うことが重要である。

コクラン・ハンドブック (Cochrane Handbook)

The Cochrane handbook for systematic reviews of intervention[6]、Cochrane のシステマティック・レビューの準備段階および維持段階に関する詳細情報が含まれている。内容は医療介入の影響ならびに一般的なメタアナリシスの資料に関する情報である。

医療研究・品質調査機構（米） Agency for Healthcare Research and Quality (AHRQ):

Methods guide for effectiveness and comparative effectiveness reviews

米国保健福祉省内部に存在する AHRQ は、情報に基づく選択およびヘルスケアサービスの質の改善に役立つ研究を支援している。AHRQ は使命の一環として、有効性比較レビュー実施のための『メソッドガイド』を出版している[291]。

Institute of Medicine (IOM): Standards for systematic reviews

米国医学研究所は、客観的、明白、かつ科学的に有効なシステマティック・レビューを保証する目的で、内科的または外科的介入の有効性比較を行うシステマティック・レビューの推奨基準を発表している[91]。

Enhancing the QUality and Transparency Of health Reseach (EQUATOR[17])

本組織は、保健研究の執筆および出版のためのリソースを収集している。本組織は報告ガイドラインの包括的な検索データベースを有し、研究報告に関連する他のリソースとリンクしている。

Centre for Reviews and Dissemination (CRD)

英国 CRD は、Systematic reviews: CRD's guidance for undertaking reviews in health care[95]というタイトルの有用な報告を作成している。

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)

PRISMA 声明および PRISMA チェックリストのうち、一部は出版されている。PRISMA チェックリストには主に、システマティック・レビューおよびメタアナリシスの報告のためのエビデンスに基づく最小限項目が含まれている[18-20、266、269]。また、PRISMA 声明にはプロトコルに含むべき内容についての助言も記載されている[267、268]。

Meta-analysis of observational studies in epidemiology (MOOSE) guidelines

MOOSE ガイドライン[265]は、観察研究のメタアナリシスおよびシステマティック・レビューに向けられている。

Annex 1 to the Guide on methodological standards in pharmacoepidemiology (European Medicines Agency)

本文書は、ENCePP Guide on methodological standards in pharmacoepidemiology の一部として、安全性アウトカムの比較薬剤疫学研究のうち完了しているもののシステマティック・レビューおよびメタアナリシスを実施するうえでのガイダンスを含んでいる (http://www.encepp.eu/standards_and_guidances/documents/ENCePP_Methods_Guide_Annex1.pdf)。

規制分野においては、当局からの具体的な要求事項に留意することが重要である。米国 FDA および EMA は現行のガイダンスをウェブサイト公開しており、これらは、計画プロセスの一部として考慮の対象とすべきものである。例えば 2015 年時点で、EMA のガイダンスとして以下が存在する EMEA/CHMP: Points to consider on application with 1. Meta-analysis and 2. one pivotal study [292]。当局への提出目的でメタアナリシスを準備する場合、計画段階の一部として、直接当局の科学的助言を求めることを考慮してもよい。

一部の出版物では、有害事象および安全性データに関するシステマティック・レビューおよびメタアナリシスの様々な側面について議論されている[6、48、60、82、86、87]。例えば Cochrane Adverse Effects Methods Group[87] および Safety Planning, Evaluation and Reporting Team (SPERT)は、有害事象および安全性データノメタアナリシスに関する詳細なガイダンスを提供している[60]。他の関連ガイダンス文書には、International Conference on Harmonisation が出版した ICH M4E[88]、米国食品医薬品局の FDA Guidance on premarketing risk assessment[89]、CIOMS Working Group VI

の報告[59]などがある。

6.3 検索戦略およびデータベース

メタアナリシスの成功にとって、優れた検索戦略が重要である。潜在的に関連がある適切な用語をすべて特定する他にも、検索するデータベースの検討が必要である。検討すべきデータベースは多数存在する。スポンサーおよび規制当局は専用の内部データベースを有するが、利用可能であれば公共アクセス可能なデータベースもレビューし、スポンサーが収集した以外の資料を獲得すべきである。しかし例外として、市販前の初期段階で、外部データが限定されている場合、当局の審査官が、内部保持データのみからメタアナリシスを実施するようスポンサーに要請する場合がある。研究機関および研究者は公的ドメインのデータベースを選択できる。一部は無料アクセス可能だが、多くはアクセスに費用が伴う。データベース検索を無料ソースに制限すると、最適なデータ収集は達成できない。また、行政関連文書からの臨床データへのアクセスを促進しようという意向が増えつつあり、一部では法的義務も存在する。

Medline および Embase は主要な国際医学文献データベースである。これらのデータベースの使用は、メタアナリシスの実施を検討する者であれば誰にとっても不可欠な出発点とみなされるべきである。

- MEDLINE は米国国立医学図書館（National Library of Medicine、NLM）の文献データベースであり、NLM Medical Subject Headings (MeSH®) により索引した 5500 冊を超える生物医学雑誌の記事 2200 万件以上を含んでいる。
www.nlm.nih.gov/pubs/factsheets/medline.html
- Elsevier Life Science Solution の Embase は生物医学的データの包括的国際データベースであり、現在出版されている 8400 以上のジャーナル誌から 2800 万件以上のデータを有する。www.elsevier.com/solutions/embase
- PubMed (MEDLINE)は、Medline、生命科学雑誌、およびオンライン書籍からの引用 2300 万件超へのアクセスを提供するウェブ検索インターフェースであり、参加発表者のウェブサイトに掲載されている全文記事へのリンクを含んでいる。PubMed は米国 National Center for Biotechnology Information (NCBI) により無料で提供されている。www.ncbi.nlm.nih.gov/pubmed

- Cochrane Library (Cochrane Central Register of Controlled Trials など) には、Cochrane Database of Systematic Reviews (CDSR) や Cochrane Central Register of Controlled Trials などの有用なリソースが多数含まれている。
www.cochranelibrary.com/
- Web of Science (Science Citation Index) は、Thomson Reuters により維持されている論文引用索引サービスである。Biosis Previews は Thomson Reuters Web of Science 研究プラットフォームの文献データベースサービスのひとつである。
<http://wokinfo.com/>
- Cumulative Index to Nursing and Allied Health Literature (CINAHL) は、広範な話題を取り扱った最上級の看護・医療関連文献の索引を提供している。文献には看護、生物医学、医学図書館学、代替/補完医療、消費者健康、および 17 の医療関連領域がある。
<https://health.ebsco.com/products/the-cinahl-database>
- PsycINFO®は行動科学および精神衛生の査読済み文献に特化した抄録・索引データベースである。
<http://www.apa.org/pubs/databases/psycinfo/index.aspx>.
- POPLINE は Johns Hopkins Bloomberg School of Public Health/Center for Communication Programs にある Knowledge for Health (K4Health) Project によって維持されている無料リソースであり、米国国際開発庁 (United States Agency for International Development、USAID) が資金提供している。POPLINE に含まれる情報には、人口および家族計画、特に避妊法研究、家族計画サービス、ヒト受精率の研究、母子健康、発展途上国における HIV/AIDS、プログラムの運用および評価、人口統計、その他関連する健康、法律、政策問題などがある。
<https://www.k4health.org/k4health-products#POPLINE>
- LILACS は中南米地域およびカリブ海地域の科学、技術文献の包括的索引である。
<http://lilacs.bvsalud.org/en/>
- African Index Medicus (AIM) は世界保健機構が Association for Health Information and Libraries in Africa (AHILA) と共同で制作した、アフリカにおける健康関連の文献および情報ソースの国際的索引である。
<http://indexmedicus.afro.who.int/>
- WHO Global Health Library (GHL)は、WHO の国際公衆衛生における知識管理戦略の一環としての取り組みである。
<http://www.who.int/library/en/>

- ProQuest の Sociological Abstracts は、社会学、ならびに社会科学および行動科学における関連領域の国際的文献の抄録・索引のデータベースである。
<http://www.proquest.com/libraries/academic/databases/sociology.html>
- ClinicalTrials.gov は、米国国立衛生研究所のサービスであり、世界中で実施されているヒト被験者に対する公費・私費双方による臨床試験の登録・結果データベースである。本ウェブサイトは元来、治験プロトコルのレポジトリとして制作されたが、現在は治験結果の要約をも含んでいる。
- The European Union Clinical Trials Registry は欧州医薬品庁 (European Medicines Agency、EMA) により主催されており、欧州連合 (EU) 加盟国および欧州経済領域 (EEA) における臨床試験、ならびに paediatric investigation plan (PIP) に参加している EU/EEA 外臨床試験の情報を含んでいる。www.clinicaltrialsregister.eu/
- ISRCTN は登録・収集データベースであり、WHO の International Clinical Trials Registry Platform (ICTRP) および International Committee of Medical Journal Editors (ICMJE) のガイドラインに基づき研究の導入説明に不可欠と考えられる基本的なデータ項目が含まれている。データベース内の研究記録はすべて無料でアクセス・検索が可能であり、記録には ISRCTN の ID が割り当てられている。ISRCTN は 2000 年、RCT のプロスペクティブ登録に賛成する声の増加に応えるかたちで立ち上げられた。元来 ISRCTN は「International Standard Randomized Controlled Trial Number」の略であったが、年月を経て ISRCTN の範囲が拡大し、ランダム化対照試験の枠を超え、ヒト集団における医療介入の有効性評価のためにデザインされたあらゆる試験が含まれるようになった。Current Controlled Trials は、Current Controlled Trials Limited が運営していたウェブサイトであり、ユーザーは RCT に関する情報を検索、登録、共有できる。現在は BioMedCentral と提携している。
www.controlled-trials.com/
- International Clinical Trials Registry Platform (ICTRP) は WHO のデータベースであり、臨床試験の自発的登録、ならびに臨床試験のデザイン、実施、および管理について国際的に合意された情報群の発表を目的としている。
<http://apps.who.int/trialsearch/>

メタアナリシスに適したデータベースの発展およびアクセスは変化し続けている。上記の提案は絶対的ではなく、時間とともに変化するものであり、固有の機会、課題、および制

限がありつつも、選択肢の可能性のレパートリーを提供するものである。上記のいずれも、必ずしもあるものが他より重要である訳ではない。メタアナリシスに使用するデータベースについて実行可能なシナリオを議論する必要があるとともに、代替策を事前に用意することを考慮する必要がある。

付録 1 事例研究—成人における抗うつ薬と自殺

本事例によって示される主要事項と問題点の要約

この例は：

- 薬剤クラスの問題の評価に関連した事項を示している；
- 予め立てた仮説や選択基準、アウトカム指標、マスター・データ・ファイルおよび分析などの後ろ向きデータの前向きな分析計画を説明している；
- 均一、後ろ向き、盲検下でのアウトカム判定の例を示している；
- わずかなアウトカムデータでの効果推定の試行による層別化などの方法の問題に対処する方法を説明している；
- サブグループや経時的なハザードパターンの検討のための患者レベルデータの利用を示している；
- 統計的有意性、一貫性のある結果の多重検定、第 1 種および第 2 種の過誤の検討を提起している；
- 場合によっては限られた情報(例 ある薬剤クラス内のいくつかの薬剤に関する情報)に基づいて規制措置が必要となることに関する問題を強調している；

1. 背景

うつ病患者は自殺リスクがかなり高い[293]。抗うつ薬はうつ病の程度を低減させる能力に基づいて販売されている。逆に、抗うつ薬と自殺との関連に関する推測は事例証拠に基づいて何十年も前から存在していた。1991 年に米国食品医薬品局 (FDA) は諮問委員会の会合を開催して自殺イベントのリスクと一般的に使用されている抗うつ薬のクラスである選択的セロトニン再取り込み阻害薬を検討した。この委員会は因果関係を証明するには証拠が不十分であると結論した。2003 年にある抗うつ薬の製造業者は抗うつ薬と小児患者の自殺関連イベントとの関連を示唆するデータを FDA に提供した。FDA は関連の検討の対象を、小児集団を対象として検討されたすべての抗うつ薬に拡大して 4,582 人の患者が含まれた 24 件の試験を対象としたメタアナリシスを実施した。自殺行動と自殺念慮からなる複合アウトカムに関しての抗うつ薬とプラセボの比較では抗うつ薬に統計的に有意な効果があった[9]。これらの試験では自殺完遂例がなかった点を重視すべきである。この結果および外部諮問委員会からの助言に基づいて、FDA はすべてのクラスの抗うつ薬の添付文書に黒枠警告を追加した。その後、FDA は成人患者を対象としたメタアナリシスを実施した[271]。この例に関する詳細は米国 FDA から公開されている説明資料一式 (briefing package) の中に記載されている[294]。

2. メタアナリシスの概要

米国 FDA のメタアナリシスでは予め書面にて定めた解析計画を使用した。この解析計画には、選択基準、仮説、アウトカムの定義、要求されている変数だけでなくコードの方法、解析手法、感度解析およびサブグループまでが明確に定義されたマスター・データ・ファイル（共通データモデル）が定められていた。

FDA は、抗うつ薬の製造業者が 20 人以上の患者を対象として実施し、完了した二重盲検ランダム化プラセボ比較試験のすべてのデータを要求した。追加要件の一つは患者レベルのデータの入手であった。薬剤の製造業者が利用可能であった治験のデータと他の利用可能な試験データをメタアナリシスの基礎にすることは、いくつかの重要な利点がある。例えば、出版が試験の組み入れに影響しないことから、治験のセットは出版バイアスの悪影響を受けない。また、規制当局からの要求により、薬剤の製造業者による試験には、薬歴やベースライン特性、患者の内訳、患者のアウトカム、有害事象などの詳細な患者レベルのデータが含まれることが多い。さらに、比較的少数の薬剤製造業者（9 社）に焦点を合わせることによって、大量の関連データのタイムリーな収集が可能となった。このメタアナリシスでは利用可能なデータがある合計 372 件の治験のデータが得られた。ほとんどの治験は実施期間が比較的短期間（12 週以下）であった。

均一な質の高いアウトカム指標を作り出すことが重要であった。評価対象のアウトカムは自殺行動および自殺念慮イベントであったが、これらは試験の主たる焦点ではなかったことから、試験において積極的かつ一貫して捕捉されていないことが多かった。FDA は薬剤製造業者に対し、イベントを後ろ向きに同定し判定するための計画に関する指示を与えた。有害事象データを用いてコード化された逐語的な有害事象用語およびコメント欄のテキスト検索を含む指定されたアルゴリズムで潜在的イベントを同定した。これらのイベントからイベントのナラティブを作成した。曝露状況に関して盲検化されたイベントのナラティブに基づいて、資格を有する担当者が自殺評価のためのコロンビア分類アルゴリズム（C-CASA）を利用してイベントを自殺完遂、自殺企図、切迫した自殺行動に向けた準備行動または自殺念慮などの特定のアウトカムに分類した[82]。このアルゴリズムは信頼性が高いことが示されていることから、全体的なプロセスで治験および患者のアウトカム指標の定義に一貫性が生まれた。

主要エンドポイントは C-CASA 分類の最初の 4 カテゴリーを占める自殺行動および自殺念慮で、一次解析対象集団は大うつ病性障害、その他のうつ病性障害およびその他の精神障害に関して実施された全ての試験であった。二次解析は他の適応群で構成された。治療中

または治療中止後 1 日以内に起こったイベントのみを解析対象とした。

一次解析は層別化された OR に関する固定効果「正確」法で、試験別に層別化した[295]。これによってプラセボとの比較による OR の推定値が得られた。このメタアナリシスで薬剤効果の推定に用いた手法は、試験内比較を維持することとイベント発現率が低い場合であっても上手く解析することの両方を目的として選択された。試験内比較を維持しないと、背景発現率や割付比が試験間で異なることから、全体的な推定値には真の効果が反映されない可能性がある。試験内比較を維持するために推定方法は試験別に層別化した。一部の統計的推定量およびそれらの不確実性指標は大規模なサンプルサイズを必要とする。この点はアウトカムイベントの発現率が低い抗うつ薬のメタアナリシスにおける懸念の一つであった。

一次解析の頑健性を検討するためにいくつかの感度解析および裏付け解析が実施された。試験内で無作為割付けされた治療群間の曝露および観察時間の差を検討して、観察されたいかなる薬剤効果もその差によって説明できないようにした。試験間の薬剤効果の不均一性を図および統計を用いて検討した。ランダム効果法の結果を一次解析の手法（固定効果法）と比較した。2 つの手法間の結果は類似していることが判明し、試験のさまざまな重み付けが全体的な結果を変化させなかったことが示唆されている。一次解析の手法では黙示的にイベントが発生していない試験を利用しなかった。これらの試験を含めることの影響を検討するために層別化した RD 推定値を用いた。この手法はイベントが発生しなかった試験を利用するものである。ここでも 2 つの手法は質的に類似した結果を示した。サブグループは一貫性および直接的な臨床的関心のために、年齢、性別、地理的位置およびケア設定に基づいて検討された。

メタアナリシスからは、抗うつ薬と自殺行動および自殺念慮の全体的な関連に統計的有意性がないことが判明し、これは小児患者を対象とした FDA のメタアナリシスの結果とは対照的である。しかし、若年成人については、関連はほぼ統計的に有意であり、メタアナリシスは患者の年齢に関連する明確なパターンを示した（図 A.1 を参照）。小児のメタアナリシスの結果を含めるとこの傾向は裏付けられた。ここで注意すべき点は、明らかに多数のサブグループが検討されたという事実があるにもかかわらず、統計的有意性はサブグループの多重検定によって調整されていなかったということである。薬剤の安全性では一般に、潜在的な有害作用が関与する可能性があることから、作用がない場合に副作用を誤って検出するエラーの可能性と作用がある場合に副作用を誤って検出しないエラーの可能性のバランスをとらなければならない。

3. 具体的対応

FDA はメタアナリシスの結果の解釈および結果に基づいた規制措置の可能性に関して FDA 諮問委員会に助言を求めた。FDA は、黒枠警告を更新して、小児患者に加えて若年成人患者の抗うつ薬に関連した自殺行動および自殺念慮のリスクを警告することを選択した。特に治療開始期の慎重な観察の必要性が強調された。これは全体的な影響が統計的に有意ではなく、どの年齢サブグループも統計的に有害な影響を示さなかったという事実にもかかわらず実施された。決定は図 A.1 に認められた年齢傾向の一貫性と小児の研究結果の一貫性に基づいて行われた。警告は、この影響は 24 歳超の患者では認められておらず、65 歳以上の患者ではリスクが低減する可能性がある、と述べている。注目すべき点は、警告に至った理由がこの影響には患者の年齢に関連して明確なパターンがあったことであり、統計的有意性ではなかったことである。

FDA は本メタアナリシスで研究対象となっていないものも含めたすべてのクラスの抗うつ薬のラベルを更新することを選択した。この決定の理論的エビデンスは、この影響が解析を行ったほとんどの薬剤に認められたことと、その機序が不明であること、どの薬剤で影響が出るかを FDA が判定できなかったことにあった。警告では、この作用が短期試験の情報に基づいたものであることにも言及していた。短期を超える使用の場合の作用に関する結論は出されなかった。

抗うつ薬や他の薬剤クラスのメタアナリシスの経験がきっかけとなって、FDA は臨床試験における自殺イベントの前向きな収集に関するドラフトガイダンス文書を作成した[81]。このガイダンス文書では、(1) 治験中にイベントが発現した患者の認知と治療の改善および (2) 自殺イベントに対する薬剤の副作用の検出と推定の改善を目的として、具体的かつ前向きな自殺イベントの収集を推進している。

欧州医薬品庁ではヒト用医薬品委員会 (CHMP) のファーマコビジランス・ワーキング・パーティー(現在のファーマコビジランスリスク評価委員会[PRAC])がレビューを実施し、FDA と同様の結論に達している。各国の規制当局には、患者およびケア提供者に対してリスクを知らせ、特に治療開始時には慎重にモニタリングを行うよう黒枠警告を追加することが推奨された。小児では、承認されている適応症以外でこれらの薬剤を使用しないことが特に強調された。

日本では、FDA から報告された成人患者のメタアナリシスの結果を受けて、すべての抗うつ薬のラベリングを更新し、24 歳以下の患者における自殺念慮および自殺企図のリスクを使用上の注意に記載した。

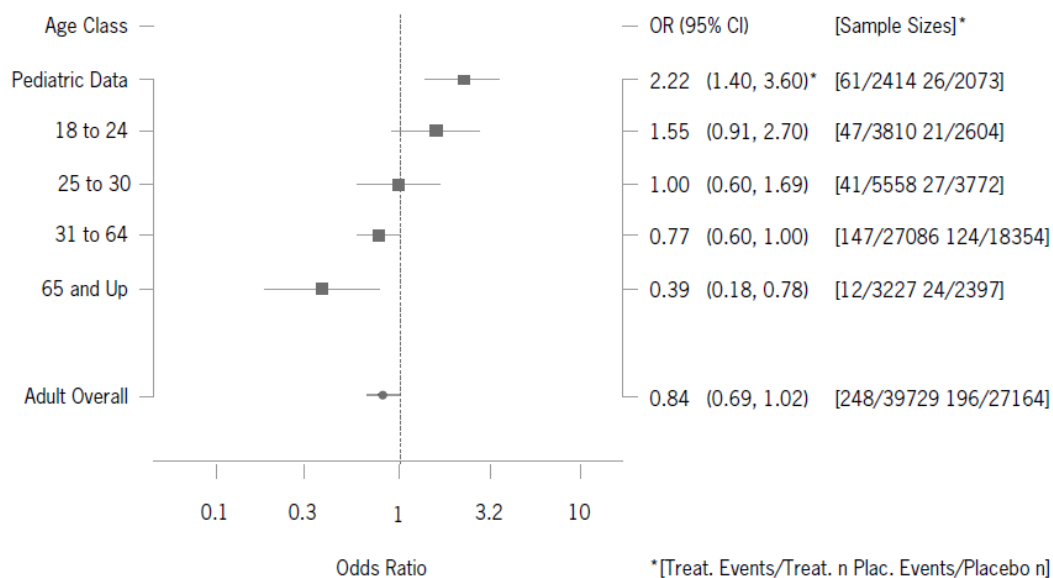


図 A.1. 自殺行動および自殺念慮に関して年齢層別のオッズ比を示した FDA のメタアナリシス（精神医学的適応症）

出典： 米国食品医薬品局（FDA）。2006年11月16日付のFDAのメモの中での言及：2006年12月13日付の精神病薬諮問委員会の概要説明文書

（www.fda.gov/ohrms/dockets/ac/06/briefing/2006-4272b1-01-FDA.pdf） この中で US FDA Statistical Evaluation of Suicidality in Adults Treated with Antidepressants（米国 FDA による抗うつ薬による治療を受けている成人の自殺傾向の統計的評価）に言及、レビュー実施日： 2006年11月17日、Levenson & Holland（米国 FDA バイオメトリクス・精神病薬部門）、38 ページの図 6、Hammad による 2004年8月16日の「Relationship between psychotropic drugs and pediatric suicidality（向精神薬と小児の自殺傾向の関連）」の再分析[81]。

付録 2 事例研究－抗腫瘍領域における赤血球造血刺激因子製剤（ESA）と血栓塞栓症

本事例で示される主要事項と問題点の要約

本事例は：

- 母集団／研究デザインの異質性は一般的であり、異質性のグローバル統計量（I² 値）が小さいからといって、治療効果の異質性に（潜在的な）特定の原因がある可能性を除外できないことに注意する；
- 無作為に起こるばらつきとメタアナリシスが個々の試験のシグナルを覆い隠すという考え方の区別が困難であることを示す；
- 検討するエンドポイントが主要（又は副次的）エンドポイントでなく有害事象として収集された時に、結果の解釈を慎重に行う必要があることを強調する（エンドポイントの重篤度が低い場合、死亡の場合と同様に慎重に検討したかどうか問う）；
- 追跡期間の異なる試験の取り扱い方を検討する；
- どの試験をメタアナリシスに組み込むべきか議論する；
- 対象とする有害事象に対してどの集団で試験を行うべきか議論する。

1. 背景

赤血球造血刺激因子製剤（ESAs）はヘモグロビン濃度を増加させ、赤血球輸血の必要性を減らし、がん患者の疲労といった貧血の症状を軽減することが出来る。しかしながら、近年いくつかの文献（Table A.2 参照）では、これらの薬剤が血栓塞栓症のリスクを増加させ、腫瘍増殖を促進する可能性があるとして報告された。これらの薬剤が生存率に影響を及ぼすかどうか、あるいはその作用機序については依然として不明である。また、慢性腎臓病患者における潜在的な副作用が確認されている。しかしながら、ここでは、不必要に複雑にしないために、単一集団に焦点を当て、貧血を伴うがん患者の研究に議論を限定する。BEST 研究（乳癌エリスロポエチン生存研究）として知られる Leyland - Jones による研究 [296] は、ESA 治療をヘモグロビン濃度が高い状態で開始し、ESAs のラベル（添付文書等の医薬品表示）で現在推奨されているよりも高いヘモグロビン濃度を目標に、1 年間、治療を継続した。この研究は ESAs が死亡率を増加させる可能性に関して、懸念を示す最初のものであった。

2. メタアナリシスの概要

これまで多くのシステマティック・レビュー及びメタアナリシスの文献が発表され、がん治療において ESA の投与で死亡率が増加するかどうかという問題に取り組んできた。これらの研究の一部を表 2 にまとめる。それぞれのメタアナリシスを詳細に記述するのではなく、これらに関連する、いくつかの特定の側面と得られた教訓を記す。

全体を通して、メタアナリシスにおける組み入れ基準として、追跡調査期間に関して絶え間ない議論が行われた。FDA の諮問委員会の会議では、「適切な生存追跡調査」の行われた研究のみを組み入れることが推奨された。これに関連して、BEST 研究では追跡調査の 4 カ月まで、その他の研究においては追跡調査の最初の数カ月以内、と早期に生存曲線が分離することに注目することが重要である。このことは、生存率に対する影響は追跡調査の早期に起こりうることを意味し、恐らく長期を意味するであろう「適切な」追跡調査という意見の重要性には疑問の余地がある。更に、懸念を生じるとした研究の中で、BEST 研究は最も大規模で、重要な最初の研究であることに注目することが重要である。確認あるいは除外したいと考えている問題のシグナルは早期の生存シグナルである。また、BEST 研究の生存曲線は追跡調査の 12 カ月後に収束するが、より長期の追跡調査の解析は、1 に近い死亡ハザード比をもたらすことに注目すべきである。このように、安全性に関して、短期間の解析は精度が高く、長期間の追跡調査では早期の生存率の差は見落とされるリスクを負う。どのような場合でも、手法を研究するメタアナリシスの専門家及び方法論者は、死亡の確認が完璧である限りは、追跡調査の期間に関わらず、通常、無作為化比較により提供された全ての情報の組み入れを推奨する。更に、死亡数の少ない研究はメタアナリシスに多くの情報を提供できず、全ての要約統計量の重みが低下する。

多くの研究において、死亡率は主要エンドポイントではなかった。これは、死亡がエンドポイントでなく単に有害事象として収集された時の、死亡の確認の完全性に疑問を投げ掛けるものである。このことは研究が有効とみなされるために、死亡のような「厳しい」エンドポイントが研究のエンドポイントである必要があるかどうかといったさらに別の疑問を生む。死亡事象はある程度努力して、研究外の情報源から確認することができる。これにより、死亡率をエンドポイントとする研究の妥当性は増加するだろう可能性が高いが、追加の追跡調査により、緊急を要する安全性の課題に対する答えに遅れを生ずる可能性がある。

結果（所見）の異質性は、実施された事実上全てのメタアナリシスにおいて重要な問題のようである。メタアナリシスにおける I^2 値は 0%、もしくは、ほぼ 0% に近かった。それにも拘わらず、変動のいくつかのパターンは、メタアナリシスを通して以下のように一貫し

ているようであった。

a) 貧血を伴うがん患者の試験では、一貫して放射線療法の試験と同じような相対リスクの増加を示す。確かに程度はメタアナリシス間で変化するが、化学療法誘発性貧血の試験に限った解析では一貫した相対リスクの低下を示す。

b) 研究レベルで調べた時（患者レベルでのベースラインのヘモグロビン濃度を階層化する Bohlius 解析は除く）[297]、ベースラインの平均ヘモグロビン濃度が低いほど相対リスクの低下を示す傾向がある。

c) 目標ヘモグロビン濃度は、ベースラインのヘモグロビン濃度よりも層別要素としての説得力が小さい。目標ヘモグロビン濃度（すなわち、ESA 治療の治療目標として定義されたヘモグロビン濃度）と達成されたヘモグロビン濃度（すなわち、治療中に得られた実際の値）を区別することは重要である。高いヘモグロビン濃度を目標とする時、早い段階で反応のない患者では、投与量が増加し、結果として達成ヘモグロビン濃度の上昇を伴う可能性がある。

d) 2006 年の AHRQ 解析[297]及び Bennett の初期バージョン（ポスターとして発表された）[298]では、「適応内」対「適応外」の層別化により「適応内」試験で相対リスクの低下が示唆された。BEST 研究を除きこれまで施行されてきた感度分析では、相対リスクは 1.0（通常約 1.03 - 1.05、信頼区間は 1.0 を跨ぐ）に近い値を示す。

3. 具体的対応

これらの薬剤の安全性はアメリカ食品医薬品局の諮問委員会会議や欧州医薬品庁の多くのヒアリング及び規制上の表示への介入の対象となり、基本的に、使用制限やリスク最小化が指示された。

表 A.2 がん治療における赤血球造血刺激因子製剤の投与により死亡率が上昇するかどうかの問題に取り組んだ公表されたシステマティック・レビュー及びメタアナリシス

筆頭著者 (発表年)	ジャーナル	試験数	被験者数	試験結果
Glaspy (2010)# [297]	Br J Cancer	60	計 15,323 人 血液療法： 12,108 人	死亡率 (60 試験) :OR=1.06 (95% CI: 0.97 - 1.15; I ² 値=0%) 化学療法のための死亡率 (47 試験) : OR=1.03 (95% CI: 0.93 - 1.13; I ² 値 =1.2%) 化学療法 (乳癌エリスロポエチン生存 試験除外) の死亡率: OR=0.98 (0.89, 1.09) VTE (44 試験) : OR=1.48 (95% CI: 1.28 - 1.72)
Bohlius (2009)* [297]	Lancet	53	計 13,933 人 化学療法： 10,441 人	ハザード比[HR]: 1.17 (95% CI: 1.06 - 1.30; I ² 値=0%) 化学療法 (38 試験) : 1.10 (0.98 - 1.24) 化学療法 (乳癌エリスロポエチン生存 試験除外) :1.03 (0.90, 1.18)
Bennett (2008) [298]	JAMA	51	13,611 人	HR: 1.10 (95% CI: 1.01 - 1.20) 化学療法:1.09 (0.99 - 1.19), I ² 値=21%
Tonelli (2009) [300]	CMAJ	52 28 (死亡率解 析) 31 (比較 試験)	計 12,006 人 6,525 人 (全死 亡率解析)	[RR]1.15 (95% CI: 1.03 - 1.29; I ² 値 =0%)

#スポンサーに所属の著者を含む。

*個人の患者レベルのデータに基づく。

略語：VTE=静脈血栓塞栓症；HR=ハザード比；OR=オッズ比；CI=信頼区間；RR=リスク比

情報源：CIOMS X 2016 により作成

付録 3. 事例研究－COPD 患者におけるチオトロピウム及び心血管イベント

本例で示される主要事項と疑問点の要約

本例では、

- ▶ 医薬品の安全性に関するメタアナリシスの発表報告には、方法論的課題がみられるものも存在する。
- ▶ 複数の小規模試験のメタアナリシスと単一の大規模試験のメタアナリシスとの間に矛盾をいかに扱うかを考察する。
- ▶ 治療を要する患者への有益な介入を不必要に否定するような治験薬の使用制限が行われないようにするため、メタアナリシスの評価にあたっては、制限について熟慮する必要があることを強調する。
- ▶ 強い影響力を有する雑誌で発表されたからと言って、自動的にエビデンスが強いとされるわけではないことを示す（特定の安全性シグナルに複数のエビデンスの流れが存在する場合、他のエビデンスの流れで確認できるまで、事後メタアナリシスによって出版された報告知見は、始めは予備的エビデンスとして扱われるべきである）。

1. 背景

慢性閉塞性肺疾患（chronic obstructive pulmonary disease: COPD）は世界における主要な死因であり、現在この疾患における肺機能低下の進行を変化させることが可能な治療薬は存在しない。米国食品医薬品局（Food and Drug Administration: FDA）は 2004 年、COPD の治療に用いられる初の長時間作用型抗コリン性気管支拡張薬として、ハンディヘラー®によるチオトロピウムの使用を承認した。チオトロピウム承認を支持する試験では、気管支拡張作用が 24 時間を超えて持続することが証明された（すなわち症状は緩和できるが、疾患変化をもたらす治療ではない）[128]。しかし、2008 年に発表されたメタアナリシスによってチオトロピウムの安全性に関する懸念が生じてきた[15]。

2. メタアナリシスの概要

2008 年 9 月、Singh らは The Journal of the American Medical Association (JAMA) において、吸入抗コリン薬に伴う心血管リスクを評価する 17 のランダム化臨床試験に対するメタアナリシスを発表した。研究報告によると、吸入抗コリン薬に対する心血管イベントの相対リスクは 1.60 (95%信頼区間: 1.22-2.10) であった (対照比較を行ったチオトロ

ピウム及びイプラトロピウムの試験を含む)。この研究では、吸入抗コリン薬の使用は心血管要因による死亡、心筋梗塞、脳卒中などの主要な有害事象リスクの有意な増加を伴うと結論された[15, 301]。

その後まもなく、New England Journal of Medicine にて Understanding Potential Long-Term Impacts on Function with Tiotropium (UPLIFT) 試験が発表されたが、この試験では以前報告されたリスクの増大は示されなかった（心血管要因による死亡においてリスク比 (risk ratio: RR) は 0.73 (95%信頼区間: 0.56-0.95) であった) [128]。UPLIFT 試験は大規模な、4 年間にわたるランダム化比較試験であり、試験を中止した患者の生命状態を含む死亡データをプロスペクティブに収集し、独立した委員会により死因の判定を行った。UPLIFT 試験における曝露群は 17,721 患者年となり、先行メタアナリシスに使用された全試験の規模を超えた[302]。

FDA の再評価では、Singh らのメタアナリシスと UPLIFT 試験との所見の差異の理由として、メタアナリシスにおける方法論的制限を強調した。

この「方法論的制限」とは、潜在的バイアスのある研究選択により心血管イベントを報告している試験に限定されたこと、患者フォローアップ時間の評価及び中止率の差異の説明がなく、プラセボ群患者の中止率が有意に高くなっていたこと（一般的にプラセボを継続投与された患者群は中止した患者群よりも健康である場合があるが、ベースラインの心血管リスク要素における潜在的差異はメタアナリシスで説明されていなかった）、メタアナリシスに含まれる試験の多くで中止となった患者における有害事象発生に関する情報が欠如していたこと、患者レベルのデータが欠如していたこと、及び短時間作用型抗コリン薬と長時間作用型抗コリン薬に対する試験が主要な分析で結合されていたことである。一方 UPLIFT 試験は、サンプルサイズが大きく、フォローアップ期間が長く、全有害事象、重篤な有害事象、あらゆる原因による死亡などの安全性エンドポイントを予め特定していた。生命状態に関する情報収集は、治療を受けた患者の 97%で 1440 日目の治療終了まで行われ、75%で 1470 日目の治療外フォローアップまで行われた。脳卒中、心血管イベント、及び死亡に関する所見は、種々の分析で一貫していた[128]。

3. 具体的対応

FDA は 2009 年 11 月 19 日、Pulmonary-Allergy Drugs Advisory Committee (PADAC) の会議を招集し、チオトロピウム曝露に伴う心血管リスク及び死亡率のデータについて議論した。PADAC は、UPLIFT のデータは脳卒中または心血管信号を適切に扱っていると結論した。FDA は、UPLIFT のデータの強度及び Singh のメタアナリシスに方法論的制

限があった可能性から、現在のエビデンスはチオトロピウムに関連して脳卒中、心臓発作、または死亡のリスクが増大するという結論を支持するものではないと結論した³⁰³⁾。

References Appendix 3

15. Singh S, Loke YK, Furberg CD. Inhaled anticholinergics and risk of major adverse cardiovascular events in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis. *JAMA*, 2008, 300(12): 1439-1450.
128. Michele TM, Pinheiro S, Iyasu S. The safety of tiotropium – the FDA’s conclusions. *N Engl J Med*, 2010, 363(12): 1097-1099.
301. Erratum J. Incorrect date in: Inhaled anticholinergics and risk of major adverse cardiovascular events in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis. *JAMA*, 2009, 301(12): 1227-1230.
302. Tashkin DP, Celli B, Senn S, Burkhart D, Kesten S, MEnjoge S, Decramer M, Investigators US. A 4-year trial of tiotropium in chronic obstructive pulmonary disease. *N Engl J Med*, 2008, 359(15): 1543-1554.
303. U.S. Food and Drug Administration. Early communication about an ongoing safety review of tiotropium [marketed as Spiriva HandiHaler]. 2008, <http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/DrugSafetyInformationforHealthcareProfessionals/ucm070651.htm>

付録 4. 事例研究—糖尿病患者におけるロシグリタゾンと

心血管イベント

本例で示される主要事項の要点

本例では、以下の内容について説明していく。

- ▶ 対象アウトカムが試験の一次的アウトカムまたは二次的アウトカムでない場合、メタアナリシスの実施にあたって課題が存在する。
- ▶ システマティック・レビューのプロトコル及び分析計画を予め指定することで、事後解析を回避できる。
- ▶ メタアナリシスにまとめて取り込まれた小規模試験の多くが、大きな重みを有することがあり、対象アウトカムを目的としてデザインされた大規模試験よりも高い重みを有することさえある（小規模試験では発生頻度の低いアウトカム研究を目的とした試験デザインが必然的に不可能であり、これがシステマティック・レビューやメタアナリシスの実施動機となる）。
- ▶ 異なる方法で異なる答えが出る場合、感度分析が重要となる。

1. 背景

II 型糖尿病、あるいはインスリン非依存性糖尿病（non-insulin dependent diabetes mellitus：NIDDM）または成人発症型糖尿病の患者では、心血管疾患（cardiovascular disease：CVD）、特に心筋梗塞（myocardial infarction：MI）のリスクが高いことが知られている。患者は明白な糖代謝異常を示すことから、血糖値の低下（血糖コントロール）が CVD、特に MI の減少につながるとこれまで推測されてきた。しかしロシグリタゾンなどの新薬が導入された時点で、CVD や MI のリスク低下を立証するエビデンスは、既存の薬剤においてすらほとんどなかった。血糖値低下による微小血管合併症（例えば失明につながり得る網膜症）の低下に対する合理的エビデンスは存在したが、MI などの大血管合併症に対する影響に関するエビデンスはほとんどなかった。

ロシグリタゾンなどの薬剤（チアゾリジンジオン系薬剤として知られる）が 1999 年に米国で初承認され、2000 年には欧州でも（第二選択薬として）初承認された。これらの新薬は、血糖コントロールのベネフィットに基づき NIDDM に対して認可された。

特筆すべきは、チアゾリジンジオン系薬剤は体液貯留増加及び体重増加を顕著に示し、CVD、特にうっ血性心不全及び MI に対するベネフィットを妨げるとされたことである。

また、少なくとも当初は欧州において、ロシグリタゾンとインスリンとの併用及びうつ血性心不全患者への使用を禁忌とされていたことも言及すべき点である。

2. メタアナリシスの概要

2007年、New England Journal of Medicine (NEJM) に Nissen と Wolski によるメタアナリシスが発表された[13]。このメタアナリシスでは116件の試験が検討され、うち42件が組み入れ基準を満たしていた。2つの大規模試験（「DREAM」及び「ADOPT」）が含まれていたが、多くは小規模試験であった。その結果、MI に対するオッズ比 (odds ratio : OR) は 1.43 (95%信頼区間 (confidence interval : CI) 1.03-1.98、 $P=0.03$)、及び心血管要素による死亡に対する OR は 1.64 (95%CI : 0.98-2.74、 $P=0.06$) となった。論文中には組み入れ基準及び除外基準の記載はなかったが、プロトコルが特定の CVD アウトカムに重点を置いたという言及はなかった。

分析には希少事象に相当とされる Peto 法（固定効果法）が用いられた。Peto 法及び類似の方法では、両群ともに事象数がゼロである試験は使用できないため、分析に寄与しない試験は MI 分析において4件、CVD による死亡の分析において19件存在した。MI 分析に用いられた試験38件のうち21件、及び CVD による死亡の分析に用いられた試験23件のうち15件では、1つの事象しか発生しなかった。事象が1つしかなかった21件のうち、16件はロシグリタゾンに対して、残り5件は対照群に対して事象が報告された。

このメタアナリシスには他にもいくつか特筆すべき特徴がある。第一に、著者らの報告では、ロシグリタゾンとインスリンの併用とインスリン単独とを比較した試験における OR は 2.78 (95%CI : 0.58-13.3、 $P=0.20$) であった。これらの試験を除外した場合、MI 及び死亡の全体的な結果は統計学的意義を失ってしまう。これは単に検出力低下に因るだけでなく（サブ分析では事象数が少ないため）、ゼロに近い OR 推定にも因る。後に削除されてはいるが、欧州でのロシグリタゾン初回承認ではインスリンとの併用が禁忌に含まれていたことは記録に備える。その他の対照薬の OR 範囲は 1.14 以上 (95%CI : 0.70-1.86、 $P=0.59$) で、メトホルミンでは 1.80 (95% CI : 0.95-3.39) までとなり、プラセボでは $P=0.07$ であった。

第二に、一部の分析では、事象数の少ない試験をまとめると、メタアナリシスにおける重みは大規模試験2つのうちいずれかの重みと同等またはそれ以上になる。小規模試験ではロシグリタゾンに対し MI 44件、CVD による死亡 25件が報告されていることからこの点は理解できるが、2つの大規模試験をまとめると、ロシグリタゾンに対し 42件の MI 及び 14件の CVD による死亡が報告されており、小規模試験がメタアナリシスにおいて実質的

に寄与していることが示されている。メタアナリシスに寄与する重みに着目した場合、OR に関しては、Peto 法での分析に対する重みへの寄与度はそれぞれ、ADOPT で 41%、DREAM で 17%、小規模試験で 42%であった。

その他の試験として MI を含む CVD の複合アウトカム検証のためにデザイン、実施された RECORD (Rosiglitazone evaluated for cardiovascular outcomes) [304]³があるが、中間分析は Nissen と Wolski の論文への回答として実施され、両氏の論文が NEJM に掲載された直後に公表された。この中間分析によると、心血管要素による入院または死亡の主要複合エンドポイントに対して、ハザード比 (hazard ratio : HR) は 1.08 で 95% CI は 0.89-1.31 であった。MI に関しては、判定された事象及び未判定の事象の双方を含む分析における HR は 1.23 (0.81-1.86) で $P=0.34$ であった。このテーマに関して FDA Advisory Committee で開かれた会議のうち 1 つ (<http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/EndocrinologicandMetabolicDrugsAdvisoryCommittee/UCM222628.pdf>) では、この研究が判定された CVD 事象の収集に特化してデザインされていたにも関わらず、FDA の科学者の一人は研究のデザイン及び実施に対しきわめて批判的であった。

Nissen と Wolski の論文 (2007) に続いて、同一のデータセットに基づくさらなるメタアナリシス及び分析が発表されたが、結論は異なるものであった (オリジナルの所見に賛同するものもあったが、MI による死亡率または CVD による死亡率のいずれかが増加していることを立証するエビデンスが不十分であると結論づけるものもあった)。一部の分析の一覧については、Böhning らの参考文献[305]を参照されたい。Nissen と Wolski は 2010 年に最新報告[306]を公表し、2007 年の論文と類似した結論を導き出している。

3. 具体的対応

2007 年、世界に向けて心血管リスクに関する警告が発せられ、体液貯留を伴ううつ血性心不全に関しては改めて警告が発せられた。RECORD 試験の報告においてもこれらの警告が強調された。

2010 年、FDA はロシグリタゾンの使用を既存患者または第二選択薬としての使用に制限した。欧州では、2010 年に欧州医薬品庁においてさらなる検討が行われ、EU 諸国におけるロシグリタゾンの使用は中止された。以降、EU においてロシグリタゾンは市販されなくなり、2016 年時点で依然利用不可能である。

2013 年末、FDA はロシグリタゾンに対する規制の多くを解除し、心血管アウトカムに特

化した試験の最終結果に対するさらなる再検討に基づいて、本薬には「標準的な II 型糖尿病薬であるメトホルミン及びスルホニル尿素と比較して、心臓発作のリスク増加が認められない」と述べた。2015 年 12 月 16 日時点で、FDA は最後まで残存していたロシグリタゾン使用制限すべてを解除した (<http://www.medscape.com/viewarticle/856056>)。

References Appendix 4

13. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med*, 2007, 356(24): 2457-2471.
304. Home PD, Pocock SJ, Beck-Nielsen H, Gomis R, Hanefeld M, Jones NP, Komajda M, McMurray JJ, RECORD Study Group. Rosiglitazone evaluated for cardiovascular outcomes-an interim analysis. *N Engl J Med*, 2007, 357(1): 28-38.
305. Böhning D, Mylona K, Kimber A. Meta-analysis of clinical trials with rare events. *Biom J*, 2015, 57(4): 633-648.
306. Nissen SE, Wolski K. Rosiglitazone revisited: an updated meta-analysis of risk for myocardial infarction and cardiovascular mortality. *Arch Intern Med*, 2010, 170(14): 1191-1201.

補遺 I. 用語集

1. Bias	バイアス.....
2. Censored/censoring	打ち切り.....
3. Channeling	チャネリング.....
4. Composite endpoint	複合エンドポイント.....
5. Confounding	交絡.....
6. Confounding by indication	処方による交絡.....
7. Covariate	共変数.....
8. Crude pooling	粗プーリング.....
9. Cumulative meta-analysis	累積メタアナリシス.....
10. Ecological bias (also known as ecological fallacy)	生態学的バイアス (生態学的誤謬ともいう)
11. Effect modifier	効果修飾要素.....
12. Endpoint	エンドポイント.....
13. Fixed effects	固定効果.....
14. Forest plot	フォレストプロット.....
15. Heterogeneity	異質性.....
16. Individual participant data	個別参加者データ.....
17. Meta-analysis	メタアナリシス.....
18. Meta-regression	メタ回帰分析.....
19. Odds ratio	オッズ比.....
20. Outcome	アウトカム.....
21. Primary endpoint	主要エンドポイント.....
22. Random effects	ランダム効果.....
23. Relative risk	相対リスク.....
24. Risk difference	リスク差.....
25. Risk ratio	リスク比.....
26. Sensitivity analysis	感度分析.....
27. Sequential meta-analysis	連続メタアナリシス.....
28. Summary-level data	要約レベルデータ.....
29. Systematic review	システマティック・レビュー.....
30. Trial	試験.....
31. Weight	重み.....

1. バイアス Bias

真実からの結果における系統的な偏り偏差。

典拠：CIOMS Xによる提案

2. 打ち切り Censored/censoring

分析実行時に対象事象が観察不能になったときに観察は打ち切られたと言う。事象が分析時に認められない場合は右（側）打ち切り、観察期間開始前に事象が認められた場合は左（側）打ち切り、観察期間内に事象の発生が記録された場合は中間打ち切りとなる。

典拠：CIOMS Xによる提案

3. チャネリング Channelling

患者アウトカムの予後予測要素の有無を踏まえ、鑑別的に薬剤を処方する状況。

典拠：Guidance for Industry and FDA Staff: "Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data," US. Food and Drug Administration, Center for Biologics and Evaluation and Research, Drug Safety, May 2013.

4. 複合エンドポイント Composite endpoint （複合エンドポイント？）

複合エンドポイントとは、複数のエンドポイントを組み合わせて単一の効果判定基準を設定したもの、各エンドポイントがそれ自体で臨床的に有意義であるものをいう。複合エンドポイントの例として、主要有害心イベント(Major Adverse Cardiac Event: MACE)がある。MACEは一般的に、心血管死亡、非致死的心筋梗塞、及び非致死の脳卒中が組み合わせたものである。

典拠：CIOMS Xによる提案

5. 交絡 Confounding

薬剤または医学的処置の使用（または回避）に影響を与え、かつ当該薬剤または処置との関連が疑われるアウトカムの確率を変える変数が存在するとき、交絡が生じる。

典拠：下記より改変（adapted from）：

Boston University School at Public Health, MPH modules, ©2016, definition of confounding.

http://sphweb.bumc.bu.edu/otlt/MPHModules/BS/BS704-EP713_Confounding&EM/

6. 処方による交絡 Confounding by indication

処方による交絡とは交絡バイアスの一種である。ある疾患の症状または徴候が治療の適応（または禁忌）に該当し、そのため薬剤または医学的処置の使用（または回避）と関連

があるとされ、結果的に薬剤の適応（または禁忌）である疾患に関連するアウトカムが高確率で生じるときに処方による交絡が生じる。

典拠：Miquel Porta, ed (2014) A Dictionary of Epidemiology (sixth ed.) Oxford University Press, ISBN-13: 9780199976737

7. 共変数 Covariate

研究のアウトカムを予測しうると考えられる変数。共変数自体が直接の研究対象である場合もあり、交絡変数または効果修飾要素が研究対象になる場合もある。

典拠：Miquel Porta, ed (2014) A Dictionary of Epidemiology (sixth ed.) Oxford University Press, ISBN-13: 9780199976737.

8. 粗プーリング Crude pooling

複数の研究データを結合し、出典を無視して単一の研究として扱う方法。

典拠：CIOMX Xによる提案

9. 累積メタアナリシス Cumulative meta-analysis

メタアナリシスのうち、研究を一つずつ特定の順番（例えば発表日または品質の順）で追加し、新たな研究が追加されるごとに結果を要約するものをいう。累積メタアナリシスのグラフでは、横軸は研究を追加するごとに要約された結果を表すものであって、単一の研究結果を表すものではない。

典拠：Glossary of Terms in The Cochrane Collaboration. _

<http://community.cochrane.org/organizational-info/resources>にて閲覧可能

10. 生態学的バイアス（生態学的誤謬ともいう） Ecological bias (also known as ecological fallacy)

集団レベルでの変数間で認められた関連性が、必ずしも個人レベルで存在する関連性を表さない、または反映しないために生じうる誤った推論。

典拠：Miquel Porta, ed(2014)A Dictionary of Epidemiology (sixth ed.) Oxford University Press, ISBN-13:9780199976737.

11. 効果修飾要素 Effect modifier

研究対象の有する特徴のうち、治療または危険要素から異なる影響を異なるレベルで受ける、すなわち特徴と治療との間に相互作用があるものをいう。この用語は通常、疫学に関連して使用される。

典拠：Dodge, Y, The Oxford Dictionary of Statistical Terms, 6th ed., International Statistical Institute, New York. Oxford University Press, Inc., 2006.

12. エンドポイント Endpoint

エンドポイント（「目標」変数、アウトカム）とは、臨床的関連のある介入効果を捕捉し、メタアナリシスにおける特定の目標に直接関連する有力な証拠が得られるように特化、設定された測定値である。

典拠：下記より適合したCIOMS Xによる提案

ICH International Conference on Harmonisation- ICH E9 Statistical principles for clinical trials ICH Harmonised Tripartite Guideline. 1995.

13. 固定効果 Fixed effects

固定効果とは、治療効果に関する個々の研究概算をメタアナリシスに結合する方法である。固定効果モデルでは、個々の研究概算における変動性はメタアナリシスに含まれない。各研究の寄与度は通常、各研究の正確性によってのみ決定される（22. ランダム効果を参照）。

典拠：CIOMS Xによる提案

14. フォレストプロット Forest plot

フォレストプロットとは、メタアナリシスに含まれる各研究の個別の結果を、結合先のメタアナリシスの結果とともにグラフに表したものである。また、フォレストプロットでは、研究結果間の異質性が読み取れるようになっている。個々の研究結果は四角で表されており、この四角は各研究の推定値に集まっている。全体的な分析における研究の重みは、推定値に描かれた四角の領域で表されることが多い。それぞれの四角を通る横軸は、各研究の信頼区間（Confidence interval: CI）を示している。常にではないが、CIは通常95%となる。メタアナリシス及びCIから得られた全体的概算は一番下に示されており、菱形で表されていることが多い。菱形の中心はプールされた推定値を表し、横の先端はCIを表す。

コクランの定義を改変したCIOMS Xによる提案（典拠： Glossary of Terms in the Cochrane Collaboration. <http://community.cochrane.org/organizational-info/resources>にて閲覧可能）

15. 異質性 Heterogeneity

異質性とは、研究間及び/または研究結果間の差異である。異質性は一般的に、臨床的異質性、方法論的異質性、及び統計的異質性の3通りに分類される²⁵²⁾。臨床的異質性とは、試験間における患者選択の差異（例えば調査対象の疾患状態、適格性基準、患者特性、または地域差）、介入の差異（例えば期間、投与量、コントロールの性質）、及びアウトカムの差異（例えばエンドポイントの定義、経過観察期間、尺度のカットオフポイン

ト)である。方法論的異質性とは、研究デザインの差異(例えばランダム化手順)及び研究実施の差異(例えば割振りの隠蔽、盲検、経過観察の中止時及び不可能時の範囲及び対応、または分析方法)である。臨床的異質性及び方法論的異質性の構成要素を決定する際には計算によらず、判断に基づいて行う。一方、統計的異質性には、個々の研究間に数値的一致がみられず、ばらつきのみを抽出して予測された値を上回る変動が生じることもあるという概念が表れている。統計的異質性は、試験間の既知の臨床的差異及び方法論的差異により生じることもあれば、未知の(臨床的及び方法論的)試験特性によって生じることも、偶然に生じることもある。

典拠：以下を適合したCIOMS Xによる提案：1) Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *BMJ*, 1994, 309(6965) and 2) Berlin JA, Crowe BJ, Whalen E, Xia HA, Koro CE, Kuebler J. Meta-analysis of clinical safety data in a drug development program: answers to frequently asked questions. *Clinical Trials*, 2013.

16. 個別参加者データ Individual participant data

参加者1名から得た複数の値を特定できるよう、研究の変数値を一覧にしたデータ。この用語は要約レベルデータと対になるが、要約レベルデータでは全結果を個別参加者データの機能として表しているため、それ以上計算しても個人に属する値を回収できなくなっている。

典拠：以下を改変したCIOMS Xによる提案：

The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Annex 1 to the Guide on Methodological Standards in Pharmacoepidemiology, 17December 2015. EMA/686352/201. http://www.encepp.eu/standards_and_guidance. にて閲覧可能

17. メタアナリシス Meta-analysis

データが複数の独立した研究から得られたものであることを適切に考慮した分析手法を用いて、共通のリサーチクエスチョンに取り組む複数の研究における定量的エビデンスの統計的な結合

典拠：CIOMS Xによる提案

18. メタ回帰分析 Meta-regression

メタアナリシスに用いられる技術であり、システマティック・レビューにおいて研究特性(例えば割付け隠蔽、ベースラインリスク、介入タイミング)と研究結果(各研究で観察された効果の規模)との間の関係を探索する方法。

典拠：Glossary of Terms in the Cochrane Collaboration.

<http://community.cochrane.org/organizational-info/resources>.にて閲覧可能

19. オッズ比 Odds ratio

ある事象の「オッズ」が「割合 / (1-割合)」であるときに、一方の「オッズ」を他方の「オッズ」で割った比。サンプルデータからオッズ比を計算する一般的な方法は、補遺II「用語集の事例研究」で説明する。

典拠：CIOMS Xによる提案

20. アウトカム Outcome

エンドポイント（12番参照）の同義語。「複合エンドポイント」（4番）も参照。

典拠：CIOMS Xによる提案

21. 主要エンドポイント Primary endpoint

主要エンドポイントとは、メタアナリシスの主目標を定義するエンドポイントまたはアウトカムである（エンドポイント、複合エンドポイント、アウトカムを参照）。

典拠：CIOMS Xによる提案

22. ランダム効果 Random effects

ランダム効果とは、治療効果に関する個々の研究概算をメタアナリシスにおいて結合する2つの方法のうち1つである。ランダム効果メタアナリシスモデルでは、個々の研究概算における変動性がメタアナリシスに含まれている。したがって、全体的概算に対する各研究の寄与度は通常、各研究の正確性及び研究間の変動性の双方から決定される。（固定効果も参照。）

典拠：CIOMS Xによる提案

23. 相対リスク Relative risk

介入または危険要素がアウトカムに対して有する影響規模の相対基準（ハザード比、オッズ比、リスク比、率比など）の総称。

典拠：CIOMS Xによる提案

24. リスク差 Risk difference

2つの割合における差異。サンプルデータからリスク差を計算する一般的な方法は補遺II「用語集の事例研究」で説明する。

典拠：CIOMS Xによる提案

25. リスク比 Risk ratio

ある割合の他方の割合に対する比。サンプルデータからリスク比を計算する一般的な方法は補遺II「用語集の事例研究」で説明する。

典拠：CIOMS Xによる提案

26. 感度分析 Sensitivity analysis

研究結果またはシステマティック・レビュー結果が、実施状況の変化にどの程度影響されるかを測定する際に用いる分析。感度分析は、使用したデータ及び方法に関する不確かな決定または仮定に対し、結果の頑強性を評価する際に用いられる。

典拠：Glossary of Terms in the Cochrane Collaboration. <http://community.cochrane.org/organizational-info/resources>.

<http://community.cochrane.org/organizational-info/resources> にて閲覧可能

27. 連続メタアナリシス Sequential meta-analysis

多重検定を行う累積メタアナリシスの特殊な形式であり、この解析においては中間解析または逐次解析に基づいて臨床試験を初期に中止（または将来の試験計画を開始前に中止）することがある。逐次解析は原則として、終了した試験で十分な証拠が得られ、これ以上の試験が不要であるかどうかを決定する際にも用いられる。

典拠：CIOMS Xによる提案

28. 要約レベルデータ Summary-level data

ある研究の参加者グループ（例えば治療群及び対照群）のレベルにおける要約統計量（例えば平均偏差、標準偏差）。

典拠：CIOMS Xによる提案

29. システマティック・レビュー Systematic review

明確に定型化された問題のレビューであり、系統的手法及び明示的手法を用いて関連する研究の特定、選択、及び批判的評価をし、レビューに含まれる研究からデータを収集、分析すること。レビューに含まれる研究結果の分析及び要約に統計的手法（メタアナリシス）を使用することもあるが、使用しないこともある。

典拠：Glossary of Terms in the Cochrane Collaboration. <http://community.cochrane.org/organizational-info/resources>.

<http://community.cochrane.org/organizational-info/resources> にて閲覧可能

30. 試験 Trial

2つ以上の介入のうち、1つを対照介入または「通常治療」とし、それらを参加者にランダムに割付けることで比較を行う実験。多くの試験では1人につき1つの介入を割り当てるが、特定の個人グループ（例えば家庭）に対し割り当てる試験、または複数の介入を個人の内（例えば異なる順番または身体の異なる部位）に割り当てる試験もある。

典拠：CIOMS Xによる提案

31. 重み Weight

治療効果についての個々の研究概算の重みとは、研究特有の概算が治療効果全体の概算に寄与する相対量である。固定効果モデルにおける重みは研究内分散の逆数により測定され、ランダム効果モデルにおける重みは研究内分散に研究間分散を加えたものの逆数により測定される。しかし、個々の研究概算の重みと固定効果モデルまたはランダム効果モデルの重みとの間に、必然的なつながりは存在しない。重みの選択は、他の考察により与えられる。

典拠：CIOMS Xによる提案

補遺 II. 用語集の事例研究－ホルモン補充療法及び乳がん における効果測定

本稿は用語集に対する補遺である。本稿では通常メタアナリシスにおいて用いられる統計尺度(指標)計算法の簡便な事例を提示し、リスク比 (RR)、リスク差 (RD)、及びオッズ比 (OR) の定義を説明する。Women's Health Initiative (WHI) 試験として知られる、ホルモン補充療法 (hormone replacement therapy : HRT) の併用とプラセボとを比較したランダム化試験が米国で実施された。主要アウトカムは浸潤性乳癌であり、結果は下表のとおりであった。表は Rossouw ら [307] が作成した表 2 からの引用である。

結果

16,000 名以上の女性被験者をランダムに割付けし、平均 5.2 年間にわたるフォローアップを実施した。乳癌発症者数及び未発症者数は下表の通りである。

投与群	乳癌発症者	乳癌未発症者	合計
HRT	166	8,340	8,506
プラセボ	124	7,978	8,102

リスクとは、ある事象（この場合は乳癌）を有する被験者数を、事象を有する可能性があった被験者数で割ったものである。ここでは、HRT 群の乳癌発症リスクは $(166/8506) = 0.0195$ 、すなわち 1.95% となる。リスクは割合または百分率で表される。一方、プラセボ群の乳癌発症リスクは $(124/8102) = 0.0153$ (1.53%) となる。

RR では通常、投与群が分子となり、対照群が分母となるため、 $RR = 0.0195/0.0153 = 1.275$ となる。RD は、 $RD = 0.0195 - 0.0153 = 0.0042$ すなわち 0.42% または 1,000 人に 4.2 人となる。

割合の代わりとなるものがオッズである。HRT 群における乳癌発症のオッズは $166/8340 = 0.0199$ である。この場合、未発症者が分母となる。プラセボ群における乳癌発症のオッズは、 $124/7978 = 0.0155$ である。

オッズはリスクよりも高いが、きわめて希少な事象の場合、オッズはリスクに限りなく近づく。OR はオッズの比であり、 $OR = 0.0199/0.0155 = 1.280$ となる。本例でも（事象が希少であるため）OR が RR に近づき、RR よりも 1 から遠ざかる。RR 及び OR が 1 であるとき、両群間に差異がないことが示唆される。

References glossary case study

252. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *BMJ*, 1994, 309(6965): 1351-1355.
307. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, Jackson RD, Beresford SA, Howard BV, Johnson KC, Kotchen JM, Ockene J, Writing Group for the Women's Health Initiative I. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principle results From the Women's Health Initiative randomized controlled trial. *JAMA*, 2002, 288(3): 321-333.

補遺 III. CIOMS X ワーキンググループメンバー及び会議

CIOMS WG X : 医薬品安全性におけるエビデンス統合とメタアナリシスは、2011年6月～2014年9月にかけて、8回の会議を開催し、2015年7月に最終編集会議を開催した。このWG報告書は6月に参加していたメンバーによってレビューされ、その後編集チームによるレビューを経て、発行に向け最終化された。編集チームのメンバーは、Brenda Crowe (Editor-in-Chief), Jesse Berlin, Tarek Hammad, Bert Leufkens, Tongtong Wang であり、Gunilla Sjölin-Forsberg, Karin Holm がサポートした。チャプター指揮は Stephen Evans (Ch.2), William Gregory (Ch. 1), Leonie Hunt (Ch. 6), George Rochester (Ch. 3、4 第一段階), Tongtong Wong (Ch. 3、4 第二段階), David Wright (Ch. 5)であった。

WG は、リスク最小化手法における一連の作業においてステークホルダーである規制当局、製薬企業、国際団体や研究機関の参加を認めた。メンバー、所属機関及びステークホルダー分類（上記で示したとおり）と以下のリストに示す。また、WG 会議の時系列一覧を示す。

名前	所属(ステークホルダー分類)	在籍期間*
Ando, Yuki	厚生労働省／PMDA、日本 (規制当局)	一部期間
Arguinzoniz, Miguel	ロッシュ (製薬企業)	全期間
Asano, Kunihito	厚生労働省／PMDA、日本 (規制当局)	一部期間
Berlin, Conny	ノバルティス (製薬企業)	全期間
Berlin, Jesse	ジョンソン&ジョンソン (製薬企業)	全期間
Boddy, Alexander	サノフィ (製薬企業)	全期間
Brueckner, Andreas	ノバルティス (製薬企業)	全期間
Cochino, Emil	EMA (規制当局)	一部期間
Cogo, Elise	Health Canada (規制当局)	一部期間
Crowe, Brenda	イーライリリー (製薬企業)	全期間
Evans, Stephen	London School of Hygiene and Tropical Medicine (研究機関)	全期間
名前	所属(ステークホルダー分類)	在籍期間*

Gaffney, Michael	ファイザー（製薬企業）	全期間
Gregory, William	ファイザー（製薬企業）	全期間
Hammad, Tarek	メルク（2014-2015）（製薬企業）、FDA、米 国（2012-2013）（規制当局）	全期間
Hobbiger, Stephen	GSK（製薬企業）	全期間
Heaton, Stephen	バイエルヘルスケア（製薬企業）	全期間
Hunt, Leonie	TGA、オーストラリア（規制当局）	全期間
Idänpään-Heikkilä, Juhana	CIOMS（シニアアドバイザー、元事務局長） （国際団体）	全期間
Irs, Alar	WHO（国際団体）	一部期間
Koh, Yvonne	HAS、シンガポール（規制当局）	一部期間
Leufkens, Bert	CBG-MEB、オランダ（規制当局）	全期間
Levenson, Mark	FDA、米国（規制当局）	一部期間
Liteplo, Robert	Health Canada（規制当局）	一部期間
Michel, Alexander	バイエルヘルスケア（製薬企業）	全期間
Quartey, George	ロッシュ（製薬企業）	全期間
Rägo, Lembit	WHO（国際団体）	一部期間
Rochester, George	FDA、米国（規制当局）	一部期間
Shah, Arvind	メルク（製薬企業）	全期間
Sjölin-Forsberg, Gunilla	CIOMS（2010-2015）（事務局長）（国際団体）	全期間
Slattery, Jim	EMA（規制当局）	全期間
Strassmann, Valerie	BfArM、ドイツ（規制当局）	全期間
Straus, Sabine	CBG-MEB、オランダ（規制当局）	一部期間
Uyama, Yoshiaki	厚生労働省／PMDA、日本（規制当局）	一部期間
Wang, Tongtong	Health Canada（2012-2014）（規制当局）	全期間
Wright, David	MHRA、英国（規制当局）	全期間
Xia, H. Amy	アムジェン（製薬企業）	全期間
Yoshida, Michihiro	武田（製薬企業）	全期間

* 「一部機関」：3年間のうち一部の機関 WG に所属したメンバー、「全期間」：すべての活動期間 WG に所属したメンバー。

CIMOS X ワーキンググループ X 会議**

日付	場所	主催
2011年6月	ジュネーブ、スイス	CIOMS
2011年11月	ジュネーブ、スイス	CIOMS
2012年3月	シルバースプリング、メリーランド、米国	FDA
2012年9月	ロンドン、英国	EMA
2013年3月	ユトレヒト、オランダ	MEB
2013年9月	シンガポール	HSA
2014年2月	ダーラム、ノースカロライナ、米国	GSK
2014年9月	バーゼル、スイス	ノバルティス
2015年7月（編集チーム）	ニューブランズウィック、ニュージャージー、 米国	ジョンソン& ジョンソン

** 旅費、宿泊費は各 WG メンバーの所属機関が負担、もしくはルールに基づき CIOMS が負担し、主催者は負担していない。

BIBLIOGRAPHICAL REFERENCES

- 1 Google. Keyword "meta-analysis" searched 26 March 2015. www.google.com.
- 2 U.S. National Center for Biotechnology Information. PubMed advanced search builder; keyword "meta-analysis" searched 26 March 2015. <http://www.ncbi.nlm.nih.gov/pubmed/advanced>.
- 3 European Medicines Agency. ICH guideline E2C (R2) on periodic benefit-risk evaluation report (PBRER). 2013, EM/VCHMP/ICH/544553/1998,1-45. http://www.ema.europa.eu/docs/en_GB/documentLibrary/Regulatory_and_procedural_guideline/2012/12/WC500136402.pdf.
- 4 Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Introduction to Meta-analysis. 2009, Chichester, UK: John Wiley & Sons.
- 5 Egger M, Smith GD, Altman D. Systematic reviews in health care: meta-analysis in context. 2nd ed. 2008, London, UK: John Wiley & Sons.
- 6 Higgins JP, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. 2011, The Cochrane Collaboration, <http://handbook.cochrane.org/>.
- 7 Khan K, Kunz R, Kleijnen J, Antes G. Systematic reviews to support evidence-based medicine. 2011. 2nd ed. 2011, London, UK: Royal Society of Medicine.
- 8 U.S. Food and Drug Administration. Guidance for industry: Diabetes mellitus-evaluating cardiovascular risk in new antidiabetic therapies to treat Type 2 diabetes. 2008. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM071627.pdf>.
- 9 Hammad TA, Laughren T, Racoosin J. Suicidality in pediatric patients treated with antidepressant drugs. *Arch Gen Psychiatry*, 2006, 63(3): 332-339.
- 10 Kim PW, Wu YT, Cooper C, Rochester G, Valappil T, Wang Y, Kornegay C, Nambiar S. Meta-analysis of a possible signal of increased mortality associated with cefepime use. *Clin Infect Dis*, 2010, 51(4): 381-389.
- 11 McMahan AW, Levenson MS, McEvoy BW, Mosholder AD, Murphy D. Age and risks of FDA-approved long-acting beta(2)-adrenergic receptor agonists. *Pediatrics*, 2011, 128(5): e1147-1154.
- 12 Colman E, Szarfman A, Wyeth J, Mosholder A, Jillapalli D, Levine J, Avigan M. An evaluation of a data mining signal for amyotrophic lateral sclerosis and statins detected in FDA's spontaneous adverse event reporting system.

- Pharmacoepidemiol Drug Saf, 2008, 17(11): 1068-1076.
- 13 Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med*, 2007, 356(24): 2457-2471.
 - 14 Nissen SE. Setting the RECORD straight. *JAMA*, 2010, 303(12): 1194-1195.
 - 15 Singh S, Lake YK, Furberg CD. Inhaled anticholinergics and risk of major adverse cardiovascular events in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis. *JAMA*, 2008. 300(12):1439-1450.
 - 16 Varas-Lorenzo C, RieraGuardia N, Calingaert B, Castellsague J, Salvo F, Nicotra F, Sturkenboom M, Perez-Gutthann S. Myocardial infarction and individual nonsteroidal anti-inflammatory drugs meta-analysis of observational studies. *Pharmacoepidemiol Drug Saf*, 2013, 22(6): 559-570.
 - 17 EQUATOR Enhancing the QUALity and Transparency Of health Research. A comprehensive searchable database of reporting guidelines, www.equator-network.org.
 - 18 Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement, *J Clin Epidemiol*, 2009, 62(10): 1006-1012.
 - 19 Stewart LA Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G, Tierney JF, Group P-ID. Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD Statement. *JAMA*, 2015, 313(16): 1657-1665.
 - 20 Mutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, Ioannidis JP, Straus S, Thoriund K, Jansen JP, Mulrow C, Catala-Lopez F, Gotsche PC, Dickersin K, Boutron 1, Altman DG, Moher D. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med*, 2015,162(11): 777-784.
 - 21 Chalmers 1, Hedges LV, Cooper H. A brief history of research synthesis. *Eval Health Prof*, 2002, 25(1): 12-37.
 - 22 Glass GV. Primary, Secondary, and Meta-Analysis of Research. *Educ Res*, 1976, 5, 3-8. <http://www.jstor.org/stable/1174772>.
 - 23 Baber NS, Lewis JA. Beta-blockers in the treatment of myocardial infarction. *BMJ*, 1980, 281(6232): 59.
 - 24 Cochrane Collaboration. Cochrane Adverse Effects Methods Group. 2007, The Group aims to develop the methods for producing high quality systematic reviews and to advise the Cochrane Collaboration on how the validity and precision of

- systematic reviews can be improved, <http://aemg.cochrane.org/>.
- 25 Berlin JA, Golub RM. Meta-analysis as evidence: building a better pyramid. *JAMA*, 2014, 312(6): 603-606.
 - 26 Eysenck HJ. Meta-analysis and its problems. *BMJ*, 1994, 309(6957): 789-792.
 - 27 Shapiro S. Meta-analysis/Shmeta-analysis. *Am J Epidemiol*, 1994,140(9): 771-778.
 - 28 Ioannidis JP, Evans SJ, Gotzsche PC, O'Neill RT, Altman DG, Schulz K, Moher D, Group C. Better reporting of harms in randomized trials; an extension of the CONSORT statement. *Ann Intern Med*, 2004,141(10): 781-788.
 - 29 Hodgkinson A, Kirkham JJ, Tudur-Smith C, Gamble C. Reporting of harms data in RCTs: a systematic review of empirical assessments against the CONSORT harms extension. *BMJ open*, 2013, 3(9): e003436.
 - 30 Kearney PM, Baigent C, Godwin J, Halls H, Emberson JR, Patrono C. Do selective cyclooxygenase-2 inhibitors and traditional non-steroidal anti-inflammatory drugs increase the risk of atherothrombosis? Meta-analysis of randomised trials. *BMJ*, 2006, 332(7553): 1302-1308.
 - 31 McGettigan P, Henry D. Cardiovascular risk and inhibition of cyclooxygenase: a systematic review of the observational studies of selective and nonselective inhibitors of cyclooxygenase 2. *JAMA*, 2006, 296(13):1633-1644.
 - 32 Hammad TA, Graham DJ, Staffa JA, Kornegay CJ, Dal Pan GJ. Onset of acute myocardial infarction after use of non-steroidal anti-inflammatory drugs. *Pharmacoepidemiol Drug Safe*, 2008,17(4): 315-321.
 - 33 Oxford University. Cancer Epidemiology Unit. 2015. <http://www.ceu.ox.ac.uk/team/valerie-beral>.
 - 34 Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. *Lancet*, 1996, 347(9017): 1713-1727.
 - 35 Beral V, Banks E, Reeves G, Wallis M. Hormone replacement therapy and high incidence of breast cancer between mammographic screens. *Lancet*, 1997, 349(9058): 1103-1104.
 - 36 Collaborative Group on Hormonal Factors in Breast C. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet*, 2001, 358(9291): 1389-1399.
 - 37 International Collaboration of Epidemiological Studies of Cervical C, Appleby P,

- Beral V, Berrington de Gonzalez A, Colin D, Franceschi S, Goodhill A, Green J, Peto J, Plummer M, Sweetland S. Cervical cancer and hormonal contraceptives: collaborative reanalysis of individual data for 16,573 women with cervical cancer and 35,509 women without cervical cancer from 24 epidemiological studies. *Lancet*, 2007, 370(9599): 1609-1621.
- 38 Lievre M, Cucherat M, Leizorovicz A. Pooling, meta-analysis, and the evaluation of drug safety. *Curr Control Trials Cardiovasc Med*, 2002, 3(1): 6.
- 39 Chuang-Stein C, Beltangady M. Reporting cumulative proportion of subjects with an adverse event based on data from multiple studies. *Pharmaceut Statist*, 2011,10(11): 3-7.
- 40 European Medicines Agency. ICH guideline E2F on development safety update report. 2010.
http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500097061.pdf,
- 41 ICH International Conference on Harmonisation. ICH E9 Statistical principles for clinical trials ICH Harmonised Tripartite Guideline. 1995.
<http://www.ich.org/products/guidelines/efficacy/efficacy-single/article/statisticalprinciples-for-clinical-trials.html>.
- 42 ISIS-4 (International Study of Infarct Survival). ISIS-4: a randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. ISIS4 (Fourth International Study of Infarct Survival) Collaborative Group. *Lancet*, 1995,345(8951):669-685.
- 43 ISIS-3 (International Study of Infarct Survival). ISIS-3: a randomised comparison of streptokinase vs tissue plasminogen activator vs anistreplase and of aspirin plus heparin vs aspirin alone among 41,299 cases of suspected acute myocardial infarction. ISIS-3 (Third International Study of Infarct Survival) Collaborative Group. *Lancet*, 1992, 339(8796): 753-770.
- 44 ISIS-2 (International Study of Infarct Survival). Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. ISIS^2 Second International Study of Infarct Survival) Collaborative Group. *Lancet*, 1988, 2(8607): 349-360.
- 45 ISIS-1 (International Study of Infarct Survival). Randomised trial of intravenous atenolol among 16 027 cases of suspected acute myocardial infarction: ISIS-1 . First International Study of Infarct Survival Collaborative Group. *Lancet*, 1986, 2(8498): 57-66.

- 46 Fourth International Study of Infarct Survival: protocol for a large simple study of the effects of oral mononitrate, of oral captopril, and of intravenous magnesium. ISIS4 collaborative group. *Am J Cardiol*, 1991, 68(14): 87D-100D.
- 47 Hammad TA, Neyarapally GA, Iyasu S, Staffa JA, Dal Pan G. The future of population-based postmarket drug risk assessment: a regulator's perspective. *Clin Pharmacol Ther*, 2013, 94(3): 349-358.
- 48 Hammad TA, Pinheiro SP, Neyarapally GA. Secondary use of randomized controlled trials to evaluate drug safety: a review of methodological considerations. *Clin Trials*, 2011, 8(5): 559-570.
- 49 Kalil AC. Is cefepime safe for clinical use? A Bayesian viewpoint. *J Antimicrob Chemother*, 2011, 66(6): 1207-1209.
- 50 Institute of Medicine. Discussion Framework for Clinical Trial Data Sharing: Guiding Principles, Elements, and Activities. 2014, Washington, DC: The National Academies Press.
- 51 Olson S, Downey AS, Institute of Medicine (U.S.). Forum on Drug Discovery Development and Translation, Institute of Medicine (U.S.). Forum on Neuroscience and Nervous System Disorders, National Cancer Policy Forum (U.S.), Institute of Medicine (U.S.). Roundtable on Translating Genomic-Based Research for Health, Institute of Medicine (U.S.). Board on Health Sciences Policy, Institute of Medicine (U.S.). Board on Health Care Services. Sharing clinical research data : workshop summary. 2013, Washington, D.C.: The National Academies Press.
- 52 Institute of Medicine. Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk. 2015.
<http://iom.nationalacademies.org/Reports/2015/Sharing-Clinical-Trial-Data.aspx>.
- 53 Wellcome Trust. ClinicalStudyDataRequest.com. 2014, The Wellcome Trust has taken responsibility for managing the review of research proposals and the operation of the Independent review panel (IRP). The Trust will also administer the IRP secretariat and appoint new panel members. www.wellcome.ac.uk.
<http://www.clinicalstudydatarequest.com/>.
- 54 Yale School of Medicine. Yale University Open Data Access (YODA) Project. 2015.
<http://yoda.yale.edu/welcome-yoda-project>.
- 55 Pharmaceutical Research and Manufacturers of America (PhRMA), PhRMA Principles For Responsible Clinical Trial Data Sharing. 2013.
<http://www.phrma.org/phrmapedia/responsible-clinical-trial-data-sharing>.
- 56 European Federation of Pharmaceutical Industries and Associations (EFPIA).

- Joint Principles for Responsible Clinical Trial Data Sharing to Benefit Patients. 2013. <http://transparency.efpia.eu/clinical-trials>.
- 57 European Medicines Agency. European Medicines Agency policy on publication of clinical data for medicinal products for human use. 2014. http://www.ema.europa.eu/docs/en_GB/document_library/Other/2014/10/WC500174796.pdf.
- 58 European Medicines Agency. Questions and answers on the European Medicines Agency policy on publication of clinical data for medicinal products for human use. 2015. http://www.ema.europa.eu/docs/en_GB/document_library/Report/2014/10/WC500174378.pdf.
- 59 CIOMS Working Group VI. Management of Safety Information from Clinical Trials. 2005, Geneva: Council for International Organizations of Medical Sciences.
- 60 Crowe BJ, Xia HA, Berlin JA, Watson DJ, Shi H, Lin SL, Kuebler J, Schriver RC, Santanello NC, Rochester G, Porter JB, Oster M, Mehrotra DV, Li Z, King EC, Harpur ES, Hall DB. Recommendations for safety planning, data collection, evaluation and reporting during drug, biologic and vaccine development: a report of the safety planning, evaluation, and reporting team. *Clin Trials*, 2009, 6(5): 430-440.
- 61 Hammad TA, Neyarapally GA. Legislative Policy [BJC1] and Science Considerations in the Era of Patient-Centeredness, Big Data, and Value. in *Benefit-Risk Assessment Methods in Medicinal Product Development: Bridging Qualitative and Quantitative Assessments*, 1st. 2016, CRC Press Taylor and Francis Group: Boca Raton, FL.
- 62 Higgins JPT, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, Savovic J, Schulz KF, Weeks L, Sterne JA, Cochrane Bias Methods G, Cochrane Statistical Methods G. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, 2011, 343; d5928.
- 63 Bero L. Editorial: Why the Cochrane risk of bias tool should include funding source as a standard item. 2013. <http://www.cochranelibrary.com/editorial/10.1002/14651858.ED000075>.
- 64 Sterne JA. Editorial: Why the Cochrane risk of bias tool should not include funding source as a standard item. 2013. <http://www.cochranelibrary.com/editorial/10.1002/14651858.ED000076>.
- 65 Al-Marzouki S, Roberts I, Evans S, Marshall T. Selective reporting in clinical trials: analysis of trial protocols accepted by The Lancet. *Lancet*, 2008, 372(9634):

201.

- 66 Dwan K, Gamble C, Williamson PR, Kirkham JJ, Reporting Bias G. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PLoS One*, 2013, 8(7); e66844.
- 67 University of York. PROSPERO, an international database of prospectively registered systematic reviews in health and social care. 2011, PROSPERO was launched by the NIHR on 22 February 2011 at U of York's Centre for Reviews and Dissemination. <http://www.crd.york.ac.uk/prospero/search.asp>.
- 68 Riley RD, Lambert PC, Staessen JA, Wang J, Gueyffier F, Thijis L, Boutritie F. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Stat Med*, 2008, 27(11): 1870-1893.
- 69 Riley RD, Steyerberg EW. Meta-analysis of a binary outcome using individual participant data and aggregate data. *Res Synth Methods*, 2010, 1(1): 2-19.
- 70 Light RJ, Pillemer DB. *Summing up: The science of reviewing research*. 1984, Cambridge, MA: Harvard University Press.
- 71 Greenland S, Lanes S, Jara M. Estimating effects from randomized trials with discontinuations: the need for intent-to-treat design and G-estimation. *Clin Trials*, 2008, 5(1): 5-13.
- 72 Toh S, Hernan MA. Causal inference from longitudinal studies with baseline randomization. *Int J Biostat*, 2008, 4(1): Article 22.
- 73 Toh S, Hernandez-Diaz S, Logan R, Robins JM, Hernan MA. Estimating absolute risks in the presence of nonadherence: an application to a follow-up study with baseline randomization. *Epidemiology*, 2010, 21(4): 528-539.
- 74 Hernan MA, Hernandez-Diaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials*, 2012, 9(1): 48-55.
- 75 Montedori A, Bonacini MI, Casazza G, Luchetta ML, Duca P, Cozzolino F, Abraha I. Modified versus standard intention-to-treat reporting: are there differences in methodological quality, sponsorship, and findings in randomized trials? A cross-sectional study. *Trials*, 2011, 12: 58.
- 76 Lewis JA, Machin D. Intention to treat-who should use ITT? *Br J Cancer*, 1993, 68(4): 647-650.
- 77 Cooper AJ, Lettis S, Chapman CL, Evans SJ, Waller PC, Shakir S, Payvandi N, Murray AB. Developing tools for the safety specification in risk management plans: lessons learned from a pilot project. *Pharmacoepidemiol Drug Saf*, 2008, 17(5): 445-454.
- 78 O'Neill RT. Assessment of safety, in *Biopharmaceutical Statistics for Drug*

- Development. 1988, Marcel Dekker.
- 79 Proschan MA, Lan KK, Wittes JT. *Statistical Methods for Monitoring Clinical Trials*. 2006, New York: Springer.
- 80 MedDRA Home Page. *Medical Dictionary for Regulatory Activities*. MedDRA MSSO (Medical Dictionary for Regulatory Activities Maintenance and Support Services Organization), 2015, 2013. <http://www.meddra.org/>.
- 81 U.S. Food and Drug Administration. *Guidance for Industry: Suicidal Ideation and Behavior: Prospective Assessment of Occurrence in Clinical Trials*, August 2012, accessed at. 2012.
<http://www.fda.gov/drugs/guidancecomplianceregulatoryinformation/guidances/ucm315156.htm>.
- 82 Posner K, Oquendo MA, Gould M, Stanley B, Davies M. Columbia Classification Algorithm of Suicide Assessment (C-CASA): classification of suicidal events in the FDA's pediatric suicidal risk analysis of antidepressants. *Am J Psychiatry*, 2007,164(7): 1035-1043.
- 83 Golder S, McIntosh HM, Lake Y. Identifying systematic reviews of the adverse effects of health care interventions. *BMC Med Res Methodol*, 2006, 6: 22.
- 84 Sherman RB, Woodcock J, Norden J, Grandinetti C, Temple RJ. New FDA regulation to improve safety reporting in clinical trials. *N Engl J Med*, 2011, 365(1): 3-5.
- 85 ICH International Conference on Harmonisation. *ICH E3 Guideline: Structure and Content of Clinical Study Reports*. 1995.
http://www.ich.org/fileadmin/Public_Web_Site/1CH_Products/Guidelines/Efficacy/E3/E3_Guideline.pdf.
- 86 Huang HY, Andrews E, Jones J, Skovron ML, Tilson H. Pitfalls in meta-analyses on adverse events reported from clinical trials, *Pharmacoepidemiol Drug Saf*, 2011, 20(10): 1014-1020.
- 87 Loke YK, Price D, Herxheimer A, Group CAEM. Systematic reviews of adverse effects: framework for a structured approach. *BMC Med Res Methodol*, 2007, 7: 32.
- 88 International Conference on Harmonisation. *ICH M4E Guideline on Enhancing the Format and Structure of Benefit-Risk Information in ICH*. 2014.
<http://www.ich.org/products/ctd/ctdsingle/article/revision-of-m4eguideline-on-enhancing-the-format-and-structureof-benefit-risk-information-in-ich.html>.
- 89 U.S. Food and Drug Administration. *Guidance for Industry: Premarketing Risk Assessment*. 2012.
<http://www.fda.gov/downloads/RegulatoryInformation/GuidancesUCM126958.pdf>.

- 90 Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomized controlled trials published in English and German. *Lancet*, 1997, 350(9074): 326-329.
- 91 Institute of Medicine. Finding What Works in Health Care: Standards for Systematic Reviews. 2011, Washington, DC: The National Academies Press.
- 92 Jones AP, Remington T, Williamson PR, Ashby D, Smyth RL. High prevalence but low impact of data extraction and reporting errors were found in Cochrane systematic reviews. *J Clin Epidemiol*, 2005, 58(7): 741-742.
- 93 Horton J, Vandermeer B, Hartling L, Tjosvold L, Klassen TP, Buscemi N. Systematic review data extraction: cross-sectional study showed that experience did not increase accuracy. *J Clin Epidemiol*, 2010, 63(3): 289-298.
- 94 Buscemi N, Hartling L, Vandermeer B, Tjosvold L, Klassen TP. Single data extraction generated more errors than double data extraction in systematic reviews. *J Clin Epidemiol*, 2006, 59(7): 697-703.
- 95 Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking reviews in health care. 2009. www.york.ac.uk/media/crd/Systematic_Reviews.pdf.
- 96 Ware JH, Vetrovec GW, Miller AB, Van Tosh A, Gaffney M, Yunis C, Arteaga C, Barer JS. Cardiovascular safety of varenicline: patient-level meta-analysis of randomized, blinded, placebo-controlled trials. *Am J Ther*, 2013, 20(3): 235-246.
- 97 Shuster JJ. Empirical vs natural weighting in random effects meta-analysis. *Stat Med*, 2010, 29(12): 1259-1265.
- 98 Berlin JA, Kim C. The use of meta-analysis in pharmacoepidemiology. in *Pharmacoepidemiology*, Fourth, S.B. (ed), Editor. 2005, John Wiley and Sons: Chichester. 681-707.
- 99 Colditz GA, Berkey CS, Mosteller F, Brewer TF, Wilson ME, Burdick E, Fineberg HV. The efficacy of bacillus Calmett-Guerin vaccination of newborns and infants in the prevention of tuberculosis: meta-analyses of the published literature. *Pediatrics*, 1995, 96(1 Pt 1): 29-35.
- 100 Colditz GA, Brewer TF, Berkey CS, Wilson ME, Burdick E, Fineberg HV, Mosteller F. Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature. *JAMA*, 1994, 271(9): 698-702.
- 101 Furukawa TA, Streiner DL, Hori S. Discrepancies among megatrials. *J Clin Epidemiol*, 2000, 53(12): 1193-1199.
- 102 LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. N

- Engl J Med, 1997, 337(8): 536-542.
- 103 U.S. Food and Drug Administration. Statistical review and evaluation antiepileptic drugs and suicidality. 2008, 45.
<http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/DrugSafetyInformationforHealthcareProfessionals/ucm070651.htm>.
- 104 Senn SJ. Overstating the evidence: double counting in meta-analysis and related problems. *BMC Med Res Methodol*, 2009, 9(10).
- 105 Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med*, 2002, 21(11): 1559-1573.
- 106 Gleser LJ, Olkin I. Stochastically dependent effect sizes, in *The handbook of research synthesis and meta-analysis*, 2nd, H. Cooper, L.V. Hedges, and J.C. Valentine, Editors. 1994, Russell Sage Foundation; New York.
- 107 Berlin JA, Colditz GA. The role of meta-analysis in the regulatory process for foods, drugs, and devices. *JAMA*, 1999, 281(9): 830-834.
- 108 Temple R. Meta-analysis and epidemiologic studies in drug development and postmarketing surveillance. *JAMA*, 1999, 281(9): 841-844.
- 109 Hernandez AV, Walker E, Ioannidis JP, Kattan MW. Challenges in meta-analysis of randomized clinical trials for rare harmful cardiovascular events: the case of rosiglitazone. *Am Heart J*, 2008, 156(1): 23-30.
- 110 Feinstein AR. Meta-analysis: statistical alchemy for the 21st century. *J Clin Epidemiol*, 1995, 48(1): 71-79.
- 111 Chalmers TC. Problems induced by meta-analyses. *Stat Med*, 1991, 10(6): 971-979; discussion 979-980.
- 112 Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials*, 1995, 16(1): 62-73.
- 113 Chalmers TC, Smith H, Blackburn B, Silverman B, Schroeder B, Reitman D, Ambroz A. A method for assessing the quality of a randomized control trial. *Control Clin Trials*, 1981, 2(1): 31-49.
- 114 Olivo SA, Macedo LG, Gadotti IC, Fuentes J, Stanton T, Magee DJ. Scales to assess the quality of randomized controlled trials: a systematic review. *Phys Ther*, 2008, 88(2): 156-175.
- 115 Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ*, 2001, 323(7303): 42-46.
- 116 Borghouts JA, Koes BW, Bouter LM. The clinical course and prognostic factors of

- non-specific neck pain: a systematic review. *Pain*, 1998, 77(1); 1-13.
- 117 Greenland S. Quality scores are useless and potentially misleading: reply to "Re: A critical look at some popular analytic methods". *Am J Epidemiol*, 1994, 140(3): 300-301.
- 118 Fabricatore AN, Wadden TA, Moore RH, Butryn ML, Gravalles EA, Erondy NE, Heymsfield SB, Nguyen AM. Attrition from randomized controlled trials of pharmacological weight loss agents: a systematic review and analysis. *Obes Rev*, 2009,10(3): 333-341,
- 119 Rosenbaum PR, Rubin DB. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *J R Stat Soc*, 1983: 212-218.
- 120 Lin DY, Psaty BM, Kronmal RA, Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics*, 1998, 54(3): 948-963,
- 121 Schneeweiss S, Glynn RJ, Tsai EH, Avorn J, Solomon DH. Adjusting for unmeasured confounders in pharmacoepidemiologic claims data using external information: the example of COX2 inhibitors and myocardial infarction. *Epidemiology*, 2005, 16(1): 17-24.
- 122 Sturmer T, Schneeweiss S, Avorn J, Glynn RJ. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. *Am J Epidemiol*, 2005, 162(3): 279-289.
- 123 Kesten S, Plautz M, Piquette CA, Habib MP, Niewoehner DE. Premature discontinuation of patients: a potential bias in COPD clinical trials. *Eur Respir J*, 2007, 30(5): 898-906.
- 124 Montgomery SA, Baldwin DS, Riley A. Antidepressant medications; a review of the evidence for drug-induced sexual dysfunction, *J Affect Disord*, 2002, 69(1-3): 119-140.
- 125 Garbe E, Suissa S. Hormone replacement therapy and acute coronary outcomes: methodological issues between randomized and observational studies. *Hum Reprod*, 2004, 19(1); 8-13.
- 126 Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *JAMA*, 2001, 285(4): 437-443.
- 127 Pitrou I, Boutron I, Ahmad N, Ravaud P. Reporting of safety results in published reports of randomized controlled trials. *Arch Intern Med*, 2009, 169(19): 1756-1761.
- 128 Michele TM, Pinheiro S, Iyasu S. The safety of tiotropium-the FDA's conclusions.

- N Engl J Med, 2010, 363(12):1097-1099.
- 129 Mallinckrodt C, Chuang-Stein C, McSorley P, Schwartz J, Archibald DG, Perahia DG, Detke MJ, Alphas L. A case study comparing a randomized withdrawal trial and a double-blind long-term trial for assessing the long-term efficacy of an antidepressant. *Pharm Stat*, 2007, 6(1): 9-22.
- 130 Perahia DG, Maina G, Thase ME, Spann ME, Wang F, Walker DJ, Detke MJ. Duloxetine in the prevention of depressive recurrences: a randomized, double-blind, placebo-controlled trial. *J Clin Psychiatry*, 2009, 70(5): 706-716.
- 131 Ghaemi SN. The failure to know what isn't known: negative publication bias with lamotrigine and a glimpse inside peer review. *Evid Based Ment Health*, 2009, 12(3): 65-68.
- 132 Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 1997, 315(7109): 629-634.
- 133 Sterne JA, Harbord RM. Funnel plots in meta-analysis. *Stata Journal*, 2004, 4:127-141.
- 134 U.S. National Institutes for Health. *ClinicalTrials.gov*. 2007, *ClinicalTrials.gov* is a registry and results database of publicly and privately supported clinical studies of human participants conducted around the world, <https://clinicaltrials.gov/>.
- 135 Yahav D, Paul M, Fraser A, Sarid N, Leibovici L. Efficacy and safety of cefepime: a systematic review and meta-analysis. *Lancet Infect Dis*, 2007, 7(5): 338-348.
- 136 Brocklebank D, Wright J, Gates C. Systematic review of clinical effectiveness of pressurised metered dose inhalers versus other hand held inhaler devices for delivering corticosteroids in asthma. *BMJ*, 2001, 323(7318): 896-900.
- 137 Olkin I, Sampson A. Comparison of meta-analysis versus analysis of variance of individual patient data. *Biometrics*, 1998, 54(1): 317-322.
- 138 Szczech LA, Berlin JA, Feldman HI. The effect of antilymphocyte induction therapy on renal allograft survival. A meta-analysis of individual patient-level data. Anti-Lymphocyte Antibody Induction Therapy Study Group. *Ann Intern Med*, 1998, 128(10): 817-826.
- 139 Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI, Anti-Lymphocyte Antibody Induction Therapy Study G. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Stat Med*, 2002, 21(3): 371-387.
- 140 Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *J Clin Epidemiol*, 2002, 55(1): 86-94.

- 141 Schmid CH, Stark PC, Berlin JA, Landais P, Lau J. Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. *J Clin Epidemiol*, 2004, 57(7): 683-697.
- 142 Vandembroucke JP. What is the best evidence for determining harms of medical treatment? *CMAJ*, 2006, 174(5):645-646.
- 143 Gagne JJ, Schneeweiss S. Comment on 'empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership'. *Stat Med*, 2013, 32(6): 1073-1074.
- 144 Rothman K, Greenland S, Lash T. *Modern Epidemiology*. 3rd Edition ed. 2008, Philadelphia, PA: Lippincott, Williams & Wilkins.
- 145 Strom BL. *Study Designs Available for Pharmacoepidemiology Studies*, in *Pharmacoepidemiology*, Fourth (4th). 2006, John Wiley & Sons.
- 146 Pearce N. Classification of epidemiological study designs. *Int J Epidemiol*, 2012, 41(2): 393-7.
- 147 Loke YK, Golder SP, Vandembroucke JP. Comprehensive evaluations of the adverse effects of drugs: importance of appropriate study selection and data sources. *Ther Adv Drug Saf*, 2011, 2(2): 59-68.
- 148 Golder S, Loke YK, Bland M. Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: methodological overview. *PLoS Med*, 2011, 8(5): e1001026.
- 149 Papanikolaou PN, Christidi GD, Ioannidis JP. Comparison of evidence on harms of medical interventions in randomized and nonrandomized studies. *CMAJ*, 2006, 174(5): 635-641.
- 150 Nayarapally GA, Hammad TA, Pinheiro SP, Iyasu S. Review of quality assessment tools for the evaluation of pharmacoepidemiological safety studies, *BMJ open*, 2012, 2(5).
- 151 ENCePP. European Network of Centres for Pharmacoepidemiology and Pharmacovigilance. 2006. <http://www.encepp.eu/encepp/studiesDatabase.jsp>.
- 152 Chavers S, Fife D, Wacholtz M, Stang P, Berlin J. Registration of Observational Studies: perspectives from an industry-based epidemiology group. *Pharmacoepidemiol Drug Saf*, 2011, 20(10):1009-1013.
- 153 von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandembroucke JP, Initiative S. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Epidemiology*, 2007, 18(6): 800-804.
- 154 Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and

- hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Collaborative Group on Hormonal Factors in Breast Cancer. *Lancet*, 1997, 350(9084): 1047-1059.
- 155 Salanti G, Dias S, Welton NJ, Ades AE, Golfinopoulos V, Kyrgiou M, Mauri D, Ioannidis JP. Evaluating novel agent effects in multiple-treatments meta-regression. *Stat Med*, 2010, 29(23): 2369-2383.
- 156 Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol*, 1997, 50(6): 683-691.
- 157 Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med*, 2002, 21(16): 2313-2324.
- 158 Lu G, Ades A. Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association*, 2006, 101(474).
- 159 Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res Synth Methods*, 2012, 3(2): 98-110.
- 160 Donegan S, Williamson P, D'Alessandro U, Tudur Smith C. Assessing key assumptions of network meta-analysis: a review of methods. *Res Synth Methods*, 2013, 4(4): 291-323.
- 161 Coxib and traditional NSAID Trialists' (CNT) Collaboration, Bhala N, Emberson J, Merhi A, Abramson S, Arber N, Baron JA, et al. Vascular and upper gastrointestinal effects of non-steroidal anti-inflammatory drugs: meta-analyses of individual participant data from randomised trials. *Lancet*, 2013, 382(9894): 769-779.
- 162 Mills EJ, Thorlund K, Ioannidis JP. Demystifying trial networks and network meta-analysis. *BMJ*, 2013, 346: f2914.
- 163 Deeks JJ, Altman DG. Effect measure for meta-analysis of trials with binary outcomes, in *Systematic reviews in health care: Meta-analysis in context*, 2nd, M. Egger, G.D. Smith, and D.G. Altman, Editors. 2008, BMJ Publishing: London, UK.
- 164 Sutton AJ, Cooper NJ, Lambert PC, Jones OR, Abrams KR, Sweeting MJ. Meta-analysis of rare and adverse event data. *Expert Rev Pharmacoecon Outcomes Res*, 2002, 2(4): 367-379.
- 165 Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med*, 2002, 21(11): 1575-600.
- 166 Engels EA, Schmid CH, Terrin N, Olkin I, Lau J. Heterogeneity and statistical

- significance in meta-analysis: an empirical study of 125 meta-analyses. *Stat Med*, 2000, 19(13): 1707-1728.
- 167 Localio AR, Margolis DJ, Berlin JA. Relative risks and confidence intervals were easily computed indirectly from multivariable logistic regression. *J Clin Epidemiol*, 2007, 60: 874-882.
- 168 Grieve AP. The number needed to treat: a useful clinical measure or a case of the Emperor's new clothes? *Pharm Stat*, 2003, 2(2): 87-102.
- 169 Smeeth L, Haines A, Ebrahim S. Numbers needed to treat derived from meta-analyses _sometimes informative, usually misleading. *BMJ*, 1999, 318(7197): 1548-1551.
- 170 Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med*, 1998, 17(24): 2815-2834.
- 171 Williamson PR, Smith CT, Hutton JL, Marson AG. Aggregate data meta-analysis with time-to-event outcomes. *Stat Med*, 2002, 21(22): 3337-3351.
- 172 Moodie PF, Nelson NA, Koch GG. A non-parametric procedure for evaluating treatment effect in the meta-analysis of survival data. *Stat Med*, 2004, 23(7): 1075-1093.
- 173 Sutton AJ, Abrams KR, Jones DR, Jones DR, Sheldon TA, Song F. *Methods for meta-analysis in medical research*. 2000, Chichester, UK: J. Wiley.
- 174 Simmonds MC, Higgins JP, Stewart LA, Tierney JF, Clarke MJ, Thompson SG. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clin Trials*, 2005, 2(3): 209-217.
- 175 Hedges LV, Olkin I. *Statistical methods for meta-analysis*. 1985, Orlando, FL: Academic Press.
- 176 Sweeting MJ, Sutton AJ, Lambert PC. Correction. *Stat Med*, 2006, 25:2700.
- 177 Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Stat Med*, 2004, 23(9): 1351-1375.
- 178 Bennett MM, Crowe BJ, Price KL, Stamey JD, Seaman JW, Jr. Comparison of bayesian and frequentist meta-analytical approaches for analyzing time to event data. *J Biopharm Stat*, 2013, 23(1): 129-145.
- 179 Bradburn MJ, Deeks JJ, Berlin JA, Russell Localio A. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Stat Med*, 2007, 26(1): 53-77.
- 180 Tian L, Cai T, Pfeffer MA, Piankov N, Cremieux PY, Wei LJ. Exact and efficient

- inference procedure for meta-analysis and its application to the analysis of independent 2x2 tables with all available data but without artificial continuity correction. *Biostatistics*, 2009, 10(2): 275-281.
- 181 Rucker G, Schwarzer G, Carpenter JR, Binder H, Schumacher M. Treatment-effect estimates adjusted for small-study effects via a limit meta-analysis. *Biostatistics*, 2011, 12(1): 122-142.
- 182 Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Prog Cardiovasc Dis*, 1985, 27(5): 335-71.
- 183 Greenland S, Salvan A. Bias in the one-step method for pooling study results, *Stat Med*, 1990, 9(3): 247-252.
- 184 Cai T, Parast L, Ryan L. Meta-analysis for rare events. *Stat Med*, 2010, 29(20): 2078-2089.
- 185 Firth D. Bias reduction of maximum likelihood estimates. *Biometrika*, 1993, 80(1): 27-38.
- 186 Heinze G, Schemper M. A solution to the problem of monotone likelihood in Cox regression. *Biometrics*, 2001, 57(1): 114-119.
- 187 *The Handbook of Research Synthesis and Meta-analysis*. 2nd ed. 2009, Russell Sage Foundation: New York, New York. 615.
- 188 Bailey KR. Inter-study differences: How should they influence the interpretation and analysis of results? *Stat Med*, 1987, 6(3): 351-358.
- 189 Peto R. Why do we need systematic overviews of randomized trials? *Stat Med*, 1987, 6(3): 233-244.
- 190 Poole C, Greenland S. Random-effects meta-analyses are not always conservative. *Am J Epidemiol*, 1999, 150(5): 469-475.
- 191 Berlin JA, Crowe BJ, Whalen E, Xia HA, Koro CE, Kuebler J. Meta-analysis of clinical trial safety data in a drug development program: answers to frequently asked questions. *Clin Trials*, 2013, 10(1): 20-31
- 192 FDA. Reviewer Guidance: Conducting a Clinical Safety Review on a New Product Application and Preparing a Report on the Review. 2005.
<http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm072974.pdf>.
- 193 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*, 1986, 7(3): 177-188.
- 194 Jackson D, Bowden J, Baker R. How does the DerSimonian and Laird procedure for random effects meta-analysis compare with its more efficient but harder to

- compute counterparts? *J Stat Plan Inference*, 2010, 140(4): 961-970.
- 195 Cornell JE, Mulrow CD, Localio R, Stack CB, Meibohm AR, Guallar E, Goodman SN. Random-effects meta-analysis of inconsistent effects: a time for change. *Ann Intern Med*, 2014, 160(4): 267-270.
- 196 IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Med Res Methodol*, 2014, 14: 25.
- 197 Cochran W, Cox G. *Experimental Designs*. 2nd ed. 1992, New York: Wiley.
- 198 Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*, 2002, 21(11): 1539-1558.
- 199 Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. *Stat Methods Med Res*, 2001, 10(4): 277-303.
- 200 Bayes TR. An essay towards solving a problem in the doctrine of chances. *Philosophical transactions of the royal society of London*, 1763: 370-418.
- 201 McGrayne SB. *The Theory That Would Not Die: How Bayes' Rule Cracked The Enigma Code, Hunted Down Russian Submarines, & Emerged Triumphant from Two Centuries of Controversy*. 2011, New Haven & London: Yale University Press.
- 202 Warn D, Thompson S, Spiegelhalter D. Bayesian random effects meta-analysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales. *Stat Med*, 2002, 21(11): 1601-1623.
- 203 Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc*, 2009, 172(1): 137-159.
- 204 Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Stat Med*, 2005, 24(15): 2401-2428.
- 205 Lee KJ, Thompson SG. Flexible parametric models for random-effects distributions. *Stat Med*, 2008, 27(3): 418-434.
- 206 Muthukumarana S, Tiwari RC. Meta-analysis using Dirichlet process. *Stat Methods Med Res*, 2012.
- 207 Mesgarpour B, Heidinger BH, Schwameis M, Kienbacher C, Walsh C, Schmitz S, Herkner H. Safety of off-label erythropoiesis stimulating agents in critically ill patients: a meta-analysis. *Intensive Care Med*, 2013, 39(11): 1896-1908.
- 208 Ohlssen D, Price KL, Xia HA, Hong H, Kerman J, Fu H, Quartey G, Heilmann CR, Ma H, Carlin BP. Guidance on the implementation and reporting of a drug safety Bayesian network meta-analysis. *Pharm Stat*, 2014, 13(1): 55-70.
- 209 Higgins JP, Spiegelhalter DJ. Being sceptical about meta-analyses: a Bayesian

- perspective on magnesium trials in myocardial infarction. *Int J Epidemiol*, 2002, 31(1): 96-104.
- 210 Askling J, Fahrbach K, Nordstrom B, Ross S, Schmid CH, Symmons D. Cancer risk with tumor necrosis factor alpha (TNF) inhibitors: meta-analysis of randomized controlled trials of adalimumab, etanercept, and infliximab using patient level data. *Pharmacoepidemiol Drug Saf*, 2011, 20(2): 119-130.
- 211 Kaizar EE, Greenhouse JB, Seltman H, Kelleher K. Do antidepressants cause suicidality in children? A Bayesian meta-analysis. *Clin Trials*, 2006, 3(2): 73-98.
- 212 Ibrahim JG, Chen MH, Xia HA, Liu T. Bayesian meta-experimental design: evaluating cardiovascular risk in new antidiabetic therapies to treat type 2 diabetes. *Biometrics*, 2012, 68(2): 578-586.
- 213 Hedges LV, Pigott TD. The power of statistical tests for moderators in meta-analysis. *Psychol Methods*, 2004, 9(4): 426-445.
- 214 Hedges LV, Pigott TD. The power of statistical tests in meta-analysis. *Psychol Methods*, 2001, 6(3): 203-217.
- 215 Hemminki E. Study of information submitted by drug companies to licensing authorities. *BMJ*, 1980, 280(6217): 833-36.
- 216 Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet*, 1991, 337(8746): 867-872.
- 217 Gotzsche PC. Reference bias in reports of drug trials, *BMJ (Clin Res Ed)*, 1987, 295(6599): 654-656.
- 218 Gotzsche PC. Multiple publication of reports of drug trials. *Eur J Clin Pharmacol*, 1989, 36(5): 429-432.
- 219 Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med*, 1987, 317(7): 426-432.
- 220 Tannock IF. False-positive results in clinical trials: multiple significance tests and the problem of unreported comparisons. *J Natl Cancer Inst*, 1996, 88(3-4): 206-207.
- 221 Heres S, Davis J, Maine K, Jetzinger E, Kissling W, Leucht S. Why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine: an exploratory analysis of head-to-head comparison studies of second-generation antipsychotics. *Am J Psychiatry*, 2006, 163(2): 185-194.
- 222 Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence biased medicine-selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ*, 2003, 326(7400): 1171-1173.
- 223 Little RJ, Rubin DB. *Statistical Analysis with Missing Data*. 2002, Oxford,

England: Wiley-Interscience.

- 224 National Institute for Health and Care Excellence (NICE). Guide to the Methods of Technology Appraisal. 2013.
<http://www.nice.org.uk/article/pmg9/chapter/Foreword>.
- 225 National Institute for Health and Care Excellence (NICE). The guidelines manual. 2012. <http://publications.nice.org.uk/the-guidelines-manual-pmg6>.
- 226 Turner NL, Dias S, Ades AE, Welton NJ. A Bayesian framework to account for uncertainty due to missing binary outcome data in pairwise meta-analysis. *Stat Med*, 2015, 34(12): 2062-2080.
- 227 Higgins JP, White IR, Wood AM. Imputation methods for missing outcome data in meta-analysis of clinical trials. *Clin Trials*, 2008, 5(3): 225-239.
- 228 Biester K, Lange S. The multiplicity problem in systematic reviews. XIII Cochrane Colloquium. 2005, Melbourne, Australia. 153.
- 229 Baum ML, Anish DS, Chalmers TC, Sacks HS, Smith H, Jr., Fagerstrom RM. A survey of clinical trials of antibiotic prophylaxis in colon surgery: evidence against further use of no-treatment controls. *N Engl J Med*, 1981, 305(14): 795-799.
- 230 Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med*, 1992, 327(4): 248-254.
- 231 Flather MD, Farkouh ME, Pogue JM, Yusuf S. Strengths and limitations of meta-analysis: larger studies may be more reliable. *Control Clin Trials*, 1997, 18(6): 568-579; discussion 661-666.
- 232 Whitehead A. A prospectively planned cumulative meta-analysis applied to a series of concurrent clinical trials. *Stat Med*, 1997, 16(24): 2901-2913.
- 233 Pogue JM, Yusuf S. Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. *Control Clin Trials*, 1997, 18(6): 580-93; discussion 661-666.
- 234 van der Tweel I, Bollen C. Sequential meta-analysis: an efficient decision-making tool. *Clin Trials*, 2010, 7(2): 136-146.
- 235 O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics*, 1979, 35(3): 549-556.
- 236 Lan KK, DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika*, 1983, 70(3): 659-663.
- 237 Kulinskaya E, Wood J. Trial sequential methods for meta-analysis. *Res Synth Methods*, 2014, 5(3): 212-220.
- 238 Imberger G, Wetterslev J, Gluud C. Trial sequential analysis has the potential to

- improve the reliability of conclusions in meta-analysis. *Contemp Clin Trials*, 2013, 36(1): 254-255.
- 239 Miladinovic B, Mhaskar R, Hozo I, Kumar A, Mahony H, Djulbegovic B. Optimal information size in trial sequential analysis of time-to-event outcomes reveals potentially inconclusive results because of the risk of random error. *J Clin Epidemiol*, 2013, 66(6): 654-659.
- 240 Higgins JPT, Whitehead A, Simmonds M. Sequential methods for random-effects meta-analysis. *Stat Med*. 2011, 30(9):903-921.
- 241 Thorlund K, Devereaux PJ, Wetterslev J, Guyatt G, Ioannidis JP, Thabane L, Gluud LL, Als-Nielsen B, Gluud C. Can trial sequential monitoring boundaries reduce spurious inferences from meta-analyses? *Int J Epidemiol* 2009, 38(1): 276-286.
- 242 Brook J, Thorlund K, Gluud C, Wetterslev J. Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses. *J Clin Epidemiol*, 2008, 61(8): 763-769.
- 243 Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. *J Clin Epidemiol*, 2008, 61(1): 64-75.
- 244 Jennison C, Turnbull BW. Meta-analyses and adaptive group sequential designs in the clinical development process. *J Biopharm Stat*, 2005, 15(4): 537-558.
- 245 Quan H, Ma Y, Zheng Y, Cho M, Lorenzato C, Hecquet C. Adaptive and repeated cumulative meta-analyses for safety signal detection during a new drug development process. *Statistical Technical Report #54 (Sanofi)*, 2011, 54.
- 246 Lan KK, Hu M, Cappelleri JC. Applying the law of iterated logarithm to cumulative meta-analysis of a continuous endpoint. *Statistica Sinica (Pfizer Inc.)*, 2003, 13:1135-1145.
- 247 Hu M, Cappelleri JC, Lan KK. Applying the law of iterated logarithm to control type I error in cumulative meta-analysis of binary outcomes. *Clin Trials*, 2007, 4(4): 329-340.
- 248 Lau J, Schmid CH, Chalmers TC. Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care. *J Clin Epidemiol*, 1995, 48(1): 45-57; discussion 59-60.
- 249 Chen MH, Ibrahim JG, Amy Xia HA, Liu T, Hennessey V. Bayesian sequential meta-analysis design in evaluating cardiovascular risk in a new antidiabetic drug development program. *Stat Med*, 2014, 33(9): 1600-1618.
- 250 Jeffreys H. An invariant form for the prior probability in estimation problems, in

- Royal Statistical Society of London. 1946. London.
- 251 Bender R, Bunce C, Clarke M, Gates S, Lange S, Pace NL, Thorlund K. Attention should be given to multiplicity issues in systematic reviews. *J Clin Epidemiol*, 2008, 61(9): 857-865.
- 252 Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *BMJ*, 1994, 309(6965): 1351-1355.
- 253 U.S. Food and Drug Administration. Guidance for industry non-inferiority clinical trials. US Department of Health and Human Services and US Food and Drug Administration, Washington, DC, 2010.
http://download.bion.com.cn/upload/201305/19233453_9245.pdf.
- 254 Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. in *Systematic Reviews in Health Care: Meta-analysis in context*, 2nd edition, M. Egger, G.D, Smith, and D.G. Altman, Editors. 2001, BMJ Publishing Group: London.
- 255 Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med*, 1999, 18(20): 2693-2708.
- 256 Gagnier JJ, Moher D, Boon H, Bombardier C, Beyene J. An empirical study using permutation-based resampling in meta-regression. *Syst Rev*, 2012, 1:18.
- 257 Morgenstern H. Uses of ecologic analysis in epidemiologic research. *Am J Public Health*, 1982, 72(12): 1336-1344.
- 258 Greenland S. Quantitative methods in the review of epidemiologic literature. *Epidemiol Rev*, 1987, 9:1-30.
- 259 Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med*, 1992, 116(1): 78-84.
- 260 Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA*, 1991, 266(1): 93-98.
- 261 Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet*, 2005, 365(9453): 82-93.
- 262 Olkin I. Re: "A critical look at some popular meta-analytic methods". *Am J Epidemiol*, 1994, 140(3): 297-299; discussion 300-301.
- 263 Egger M, Smith GD, Sterne JA. Uses and abuses of meta-analysis. *Clin Med*, 2001, 1(6): 478-484.
- 264 Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. *BMJ*, 2001, 322(7300):1479-1480.

- 265 Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-analysis of observational studies in epidemiology; a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA*, 2000, 283(15): 2008-2012.
- 266 Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med*, 2009, 6(7): e1000100.
- 267 Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*, 2015, 4:1.
- 268 Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA, Group P-P. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ*, 2015, 349: g7647.
- 269 Zorzela L, Lake YK, Ioannidis JP, Colder S, Santaguida P, Altman DG, Moher D, Vohra S. PRISMA harms checklist: improving harms reporting in systematic reviews. *BMJ*, 2016, DOI: <http://dx.doi.org/10.1136/bmj.il57>.
<http://www.bmj.com/content/352/bmj.il57>.
- 270 Hammad TA, Neyarapally GA, Pinheiro SP, Iyasu S, Rochester G, Dal Pan G. Reporting of meta-analyses of randomized controlled trials with a focus on drug safety: an empirical assessment. *Clin Trials*, 2013, 10(3): 389-397.
- 271 Stone M, Laughren T, Jones ML, Levenson M, Holland PC, Hughes A, Hammad TA, Temple R, Rochester G. Risk of suicidality in clinical trials of antidepressants in adults: analysis of proprietary data submitted to US Food and Drug Administration. *BMJ*, 2009, 339:b2880.
- 272 Lang T, Secic M. Considering "prior probabilities:" reporting Bayesian statistical analyses, in *How to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors, and Reviewers*. 1997, American College of Physicians: Philadelphia. 231-235.
- 273 BaSiS Group. Bayesian Standards in Science: Standards for Reporting of Bayesian Analyses in the Scientific Literature. 2001.
<http://www.stat.cmu.edu/bayesworkshop/2001/BaSiS.html>,
- 274 Sung L, Hayden J, Greenberg ML, Koren G, Feldman BM, Tomlinson GA. Seven items were identified for inclusion when reporting a Bayesian analysis of a clinical study. *J Clin Epidemiol*, 2005, 58(3): 261-268.

- 275 Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian approaches to clinical trials and health-care evaluation. 2004: Wiley. com.
- 276 Herbst AL, Ulfelder H, Poskanzer DC. Adenocarcinoma of the vagina. Association of maternal stilbestrol therapy with tumor appearance in young women. *N Engl J Med*, 1971, 284(15): 878-881.
- 277 Crowe B, Xia HA, Nilsson ME, Shahin S, Wang WV, Jiang Q. The program safety analysis plan: An implementation guide, in *Quantitative evaluation of safety in drug development: Design, analysis, and reporting*, Q. Jiang and H.A. Xia, Editors. 2015, Chapman & Hall: London. 55-68.
- 278 Xia HA, Jiang Q. Statistical evaluation of drug safety data. *Therapeutic Innovation and Regulatory Science*, 2014, 48(1): 109-120.
- 279 Ioannidis JP. Why most published research findings are false. *PLoS Med*, 2005, 2(8): e124.
- 280 Nüesch E, Trelle S, Reichenbach S, Rutjes AW, Tschannen B, Altman DG, Egger M, Juni P. Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study, *BMJ*, 2010, 341: c3515.
- 281 Moreno SG, Sutton AJ, Thompson JR, Ades AE, Abrams KR, Cooper NJ, A generalized weighting regression-derived meta-analysis estimator robust to small-study effects and heterogeneity. *Stat Med*, 2012, 31(14):1407-1417.
- 282 Kirkham JJ, Altman DG, Williamson PR. Bias due to changes in specified outcomes during the systematic review process. *PLoS One*, 2010, 5(3): e9810.
- 283 Trikalinos TA, Churchill R, Ferri M, Leucht S, Tuunainen A, Wahlbeck K, Ioannidis JP, project E-P. Effect sizes in cumulative meta-analyses of mental health randomized trials evolved over time. *J Clin Epidemiol*, 57(11): 1124-1130.
- 284 CIOMS Working Group IX. Practical approaches to risk minimisation for medicinal products: Report of CIOMS Working Group IX. 2014, Geneva, Switzerland: Council for International Organizations of Medical Sciences (CIOMS).
- 285 European Medicines Agency. Guideline on good pharmacovigilance practices (GVP): Module VI – Management and reporting of adverse reactions to medicinal products (Rev 1) [online]. 2014, 28. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2014/09/WC500172402.pdf.
- 286 Rezaie A. Absolute Versus Relative Risk: Can We Persuaded by Information Framing? *Asian J Epidemiol*, 2012. 5: 62-65.
- 287 European Medicines Agency. Guideline on good pharmacovigilance practices

- (GVP). Module XV – Safety communication. 2012.
http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/01/WC500137666.pdf.
- 288 Uppsala Monitoring Centre. Dialogue in pharmacovigilance. 2002.
- 289 Uppsala Monitoring Centre. Effective communication in pharmacovigilance: the Erice report., in International Conference on Developing Effective Communications in Pharmacovigilance. 1997, Uppsala Monitoring Centre: Erice, Sicily, Italy.
- 290 Pamer CA, Hammad TA, Wu TT, Kaplan S, Rochester G, Governale L, Mosholder AD. Changes in US antidepressant and antipsychotic prescription patterns during a period of FDA actions. *Pharmacoepidemiol Drug Saf*, 2010,19(2): 158-174.
- 291 AHRQ Agency for Healthcare Research and Quality. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. 2014.
<http://effectivehealthcare.ahrq.gov/ehec/products/60/318/CER-Methods-Guide-140109.pdf>.
- 292 European Medicines Agency. Points to consider on application with 1. meta-analyses; 2. one pivotal study. 2001.
http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003657.pdf.
- 293 Sadock BJ, Shaddock VA, Ruiz P, Kaplan HI. Kaplan & Sadock's comprehensive textbook of psychiatry. 9th ed. 2009, Philadelphia: Walters Kluwer Health/Lippincott Williams & Wilkins.
- 294 U.S. Food and Drug Administration. FDA briefing document for the December 13, 2007 Psychopharmacologic Drugs Advisory Committee meeting. 2007.
<http://www.fda.gov/ohrms/dockets/ac/06/briefing/2006-4272blindex.htm>.
- 295 Gart JL. Point and interval estimation of the common odds ratio in the combination of 2 x 2 tables with fixed marginals. *Biometrika*, 1970, 57: 471-475.
- 296 Leyland-Jones B, BEST Investigators and Study Group. Breast cancer trial with erythropoietin terminated unexpectedly. *Lancet Oncol*, 2003, 4(8): 459-460.
- 297 Bohlius J, Wilson J, Seidenfeld J, Piper M, Schwarzer G, Sandercock J, Trelle S, Weingart O, Bayliss S, Brunskill S, Djulbegovic B, Bennett CL, Langensiepen S, Hyde C, Engert E. Erythropoietin or darbepoetin for patients with cancer. *Cochrane Database Syst Rev*, 2006(3): CD003407.
- 298 Bennett CL, Silver SM, Djulbegovic B, Samaras AT, Blau CA, Gleason KJ, Barnato SE, Elverman KM, Courtney DM, McKoy JM, Edwards BJ, Tighe CC, Raisch DW, Yarnold PR, Dorr DA, Kuzel TM, Tallman MS, Trifilio SM, West DP,

- Lai SY, Henke M. Venous thromboembolism and mortality associated with recombinant erythropoietin and darbepoetin administration for the treatment of cancer-associated anemia. *JAMA*, 2008, 299(8): 914-924.
- 299 Glaspy J, Crawford J, Vansteenkiste J, Henry D, Rao S, Bowers P, Berlin JA, Tomita D, Bridges K, Ludwig H. Erythropoiesis-stimulating agents in oncology: a study-level meta-analysis of survival and other safety outcomes. *Br J Cancer*, 2010, 102(2): 301-315.
- 300 Tonelli M, Hemmelgarn B, Reiman T, Manns B, Reaume MN, Lloyd A, Wiebe N, Klarenbach S. Benefits and harms of erythropoiesis-stimulating agents for anemia related to cancer: a meta-analysis. *CMAJ*, 2009, 180(11): E62-71.
- 301 Erratum J. Incorrect data in: Inhaled anticholinergics and risk of major adverse cardiovascular events in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis. *JAMA*, 2009, 301(12): 1227-1230.
- 302 Tashkin DP, Celli B, Senn S, Burkhart D, Kesten S, Menjoge S, Decramer M, Investigators US. A 4-year trial of tiotropium in chronic obstructive pulmonary disease. *N Engl J Med*, 2008, 359(15): 1543-1554.
- 303 U.S. Food and Drug Administration. Early communication about an ongoing safety review of tiotropium [marketed as Spiriva HandiHaler]. 2008. <http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/DrugSafetyInformationforHealthcareProfessionals/ucm070651.htm>.
- 304 Home PD, Pocock SJ, Beck-Nielsen H, Gomis R, Hanefeld M, Jones NP, Komajda M, McMurray JJ, RECORD Study Group. Rosiglitazone evaluated for cardiovascular outcomes-an interim analysis. *N Engl J Med*, 2007, 357(1): 28-38.
- 305 Böhning D, Mylona K, Kimber A. Meta-analysis of clinical trials with rare events. *Biom J*, 2015, 57(4): 633-648.
- 306 Nissen SE, Wolski K. Rosiglitazone revisited: an updated meta-analysis of risk for myocardial infarction and cardiovascular mortality. *Arch Intern Med*, 2010, 170(14): 1191-1201.
- 307 Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, Jackson RD, Beresford SA, Howard BV, Johnson KG, Kotchen JM, Ockene J, Writing Group for the Women's Health Initiative I. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA*, 2002, 288(3): 321-333.