

メタアナリシス

医学統計セミナー2017 第4回目

下川敏雄
医療データサイエンス学
(附属病院臨床研究センター)

医学統計セミナー2017

11月28日(火)	診断・検査データの解析：ROC曲線を中心に
12月22日(金)	多変量解析：重回帰・ロジスティック回帰・比例ハザードモデル
1月26日(金)	傾向スコアによる解析：マッチングによる解析を中心に
2月23日(金)	メタアナリシス：2値・量的・生存期間データに対するメタ回帰分析
3月23日(金)	臨床試験における必要症例数の計算

システマティックレビューとメタアナリシス(1/2)

システマティックレビュー(Systematic review)

IOM(Institute Of Medicine：米国医学研究所)の定義

「特定の問題に絞って、類似したしかし別々の個々の問題の知見を見つけ出し、選択し、評価し、まとめるために、明確で計画された科学的方法を用いる科学研究。別々の研究からの結果の定量的統合(メタアナリシスを含むことも含まないこともある)」

システマティックレビューの条件(日本医療機能評価機構, 2017)

1. 参照した研究に漏れがない
2. 採択された研究に偏りがない
3. 中立の立場で一定の基準に基づき各研究を評価
 - ① アウトカムに及ぼす効果の大きさ
 - ② 効果の確実性
4. 結論に評価の結果が反映されている

システマティックレビューとメタアナリシス(2/2)

定量的システマティックレビュー(メタアナリシス)

定量的システマティックレビューでも、いわゆるメタアナリシスの前にバイアスの評価など定性的な評価を行い、定量的に統合できるか異質性(heterogeneity)を検討する必要がある。メタアナリシスとは、効果指標の値を統計学的に統合し、統合値と信頼区間を計算し、定量的統合を行うことである。

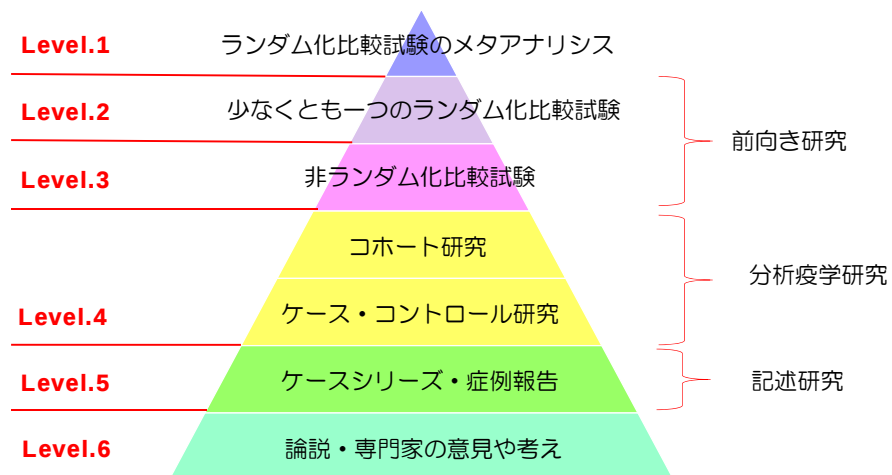
定性的システマティックレビュー

研究や除外された研究の数、対象者の特性と人数、比較と介入の方法、バイアスリスクの評価などを記述し、深い理解を与えるために定性的にまとめたものをいう。定性的システマティックレビューの結果はエビデンスの強さの判定に反映させる。

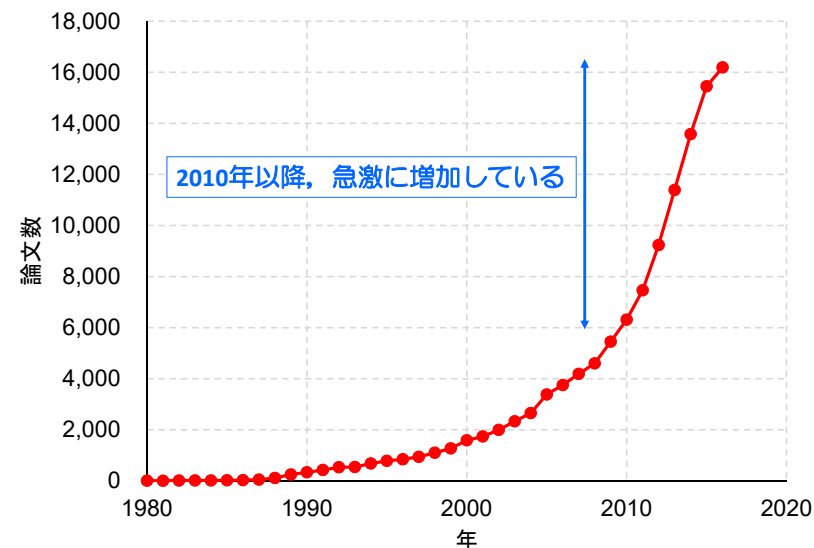
広義には、システマティックレビューもメタアナリシスと同義であるが、狭義には、システマティックレビューの一部、すなわち個々の論文の結果を統合するための統計的方法がメタアナリシスである。

研究デザインとエビデンス

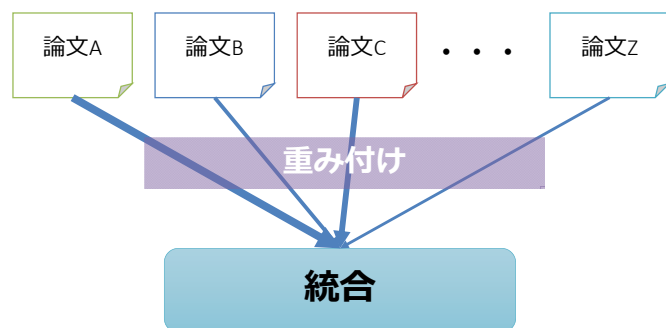
治療に関する論文のエビデンスレベル (AHRQ:米国医療政策研究局)



PubMedに掲載されている論文の推移(1980-2016)

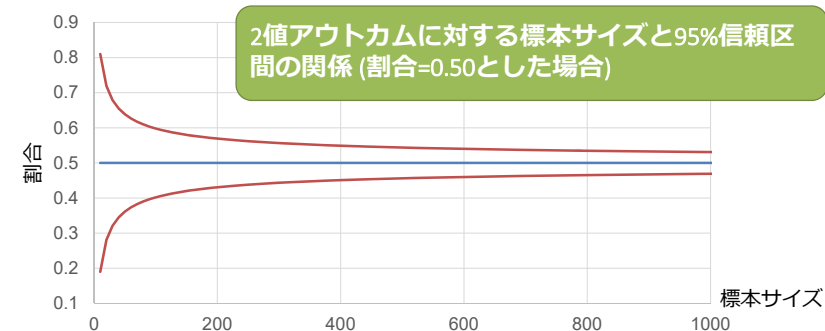


メタアナリシスとは



複数の独立した研究の結果を統計的に統合するための技法が**メタアナリシス**である。

精度の確保と症例数の関係



研究を統合することで、研究全体としての標本サイズが増える
← 精度(いいかえれば推定値に対する標準誤差)が上昇



エビデンスが高い研究結果となり得る

医学系研究における3つの目標

■ 精度の確保

精度を高めることで、明白な結論が出るように工夫する
＝ 症例数(N)を増やす(臨床試験では統計的に設定)

メタアナリシス

複数の試験を統合するのでNが増加

■ 比較可能性の保証

無作為化比較試験(RCT)では、ランダム割付を実施することで内的妥当性(internal validity)を担保している。

複数のRCTを統合しているので担保されている

■ 一般可能性(外的妥当性)の検討

単一の試験では、限られた対象患者のみになるため、一般可能性は、推測の域を脱し得ない。また、適格基準(eligibility criteria)に依存して、一般化可能性の範囲が変化する。

一貫性・交互作用の枠組みで統計学的に評価される。

メタアナリシスの手順

1. 試験の存在確認と選択

文献データベースからキーワードを入力して関連する文献を抽出する。

- ・ 事前にNGC(National Guideline Clearinghouse), NICE(National Institute for Health and Care Excellence), Mindsガイドライン等で先行するガイドラインを検討する。

- ・ MEDLINE/Pubmed, The Cochrane Library, 医中誌などを検索して、本研究に該当する試験を検索する。

→ 検索では、PICOのP(Patient), I(Intervention), C(Comparison, 場合によって)を用いて網羅的に検索する。また、複数の文献データベースを探索する。

2. データの収集と情報の抽出

文献を精査し、本研究に適合する論文を選別する。

一次スクリーニング：タイトル、アブストラクトからクリニカルクエスチョンにあてはまらないもの、採用基準に合っていないものを除外する。

二次スクリーニング：フルテキストを読み、選択基準に合った論文を選択する。選択した論文は、一覧表としてまとめる(後述)。

精査した論文は、PRISMA Flow Diagramにまとめる。

3. 解析と報告

統計的方法を利用して、効果量を推定するとともに、異質性(heterogeneity)を評価する(後述)。

論文の検索

■ データベース：MEDLINE/Pubmed, The Cochrane Library, 医中誌Web, EMBASSE 等

- ・ 単一のデータベースだけで実施しない(複数のデータベースで実施する)。
- ・ 英語論文だけでなく、日本語論文も対象にすることが推奨(英語バイアス)。
- ・ 複数名で検索を実施する。
- ・ 検索のトレーサビリティを確保する。

■ 論文検索整理の例示(診療ガイドライン作成マニュアル2017)

検索情報の記録

タイトル
クリニカルクエスチョン
データベース名
日付
検索者

No	検索式	文献数
1		
2		
3		
...		

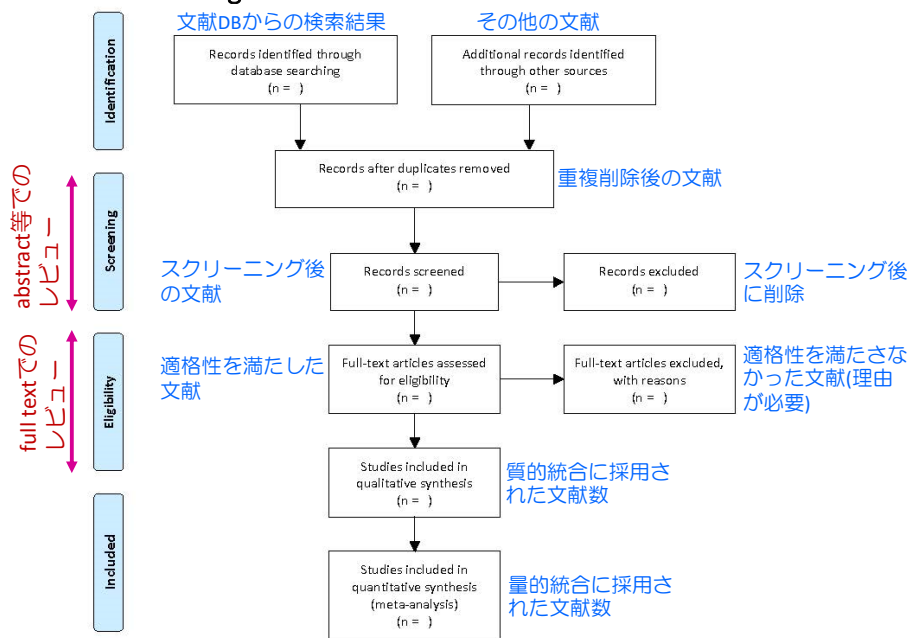
定性的レビューの手順

一次スクリーニングでは、以下の作業を行う。

- (1) 重複論文を削除する。
- (2) タイトル及びアブストラクトから、本メタアナリシスに関連しない(本研究の被験者(P), 介入(I), 場合によっては比較対象(C)に一致しない)研究を削除する。なお、これらが不明な論文は、二次スクリーニングで論文本体を精査するため、削除しないほうが良い。
- (3) 論文本体が入手できない場合には削除する。
- (4) 同一研究で異なる論文が投稿された場合(2重投稿ではなく、例えば、癌臨床試験での主論文とサブグループ解析などのように、複数の論文にまたがっているもの)は、本メタアナリシスとの関連が薄いと考えられる論文を削除する。

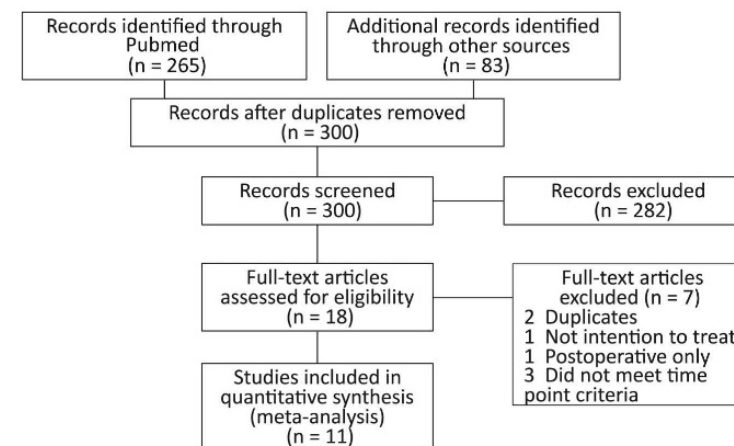
(6)その他のバイアス

PRISMA Flow diagram



PRISMA Flow diagramの例示

非心臓手術前にβ遮断薬を開始した無作為化比較試験を統合したメタアナリシス



Bouri, S. et al.: Meta-analysis of secure randomized controlled trials of β-blockade to prevent perioperative death in non-cardiac surgery, Herat, 100, 456-464, ...

統合解析：単純な意味 (丹後, 2016)

■ 平均値に関する単純な意味

(各研究の測定値) = (共通な真値) + (誤差)

■ 割合に関する単純な意味

(各研究の測定値) = (共通な割合) + (誤差)

最も単純なのは、すべての研究の平均値をとることである

➡ 平均値をとるとは、すべての研究が同じ重みであると考えることを意味する。

➡ 標準誤差(研究の精度)は標本サイズが大きくなるほど高い。つまり、標本サイズが異なる研究を等しく統合することは、異なる精度の研究結果を同じ結果として扱ってしまうことを意味する。

研究結果を平均で扱うことの危険性

	既存薬			新薬		
	標本サイズ	有効例数	有効割合 (%)	標本サイズ	有効例数	有効割合 (%)
臨床試験A	20	15	75.0	24	6	25.0
臨床試験B	240	50	20.8	230	100	43.5
臨床試験C	120	32	26.7	130	40	30.8

平均値 = 40.8

平均値 = 33.1

単純平均では、既存薬のほうが新薬よりも有効割合が高い

➡ 症例数が最も少ない臨床試験A(既存薬20例, 新薬24例)の試験結果の影響が強い

➡ 精度が最も低い研究結果が多大な影響を及ぼしている。

臨床試験の結果を単純に平均してはいけない

研究結果をまとめて扱うことの危険性

■ Simpsonのパラドックス(丹後,2016)

	新薬		標準薬	
	有効 (%割合)	症例数	有効 (%割合)	症例数
試験 1	25(83.3)	30	44(78.6)	56
試験 2	31(57.4)	54	17(54.8)	31
試験 1+2	56 (66.7)	84	61 (70.1)	87

個別の試験では、新薬のほうが標準薬よりも有効割合が高いにも関わらず、まとめて取り扱ってしまうと、結果が逆になってしまう。

➡ 同じ対象、同じ治療法であっても、それぞれの群の患者特性、医師群の特性、診療環境が異なる。

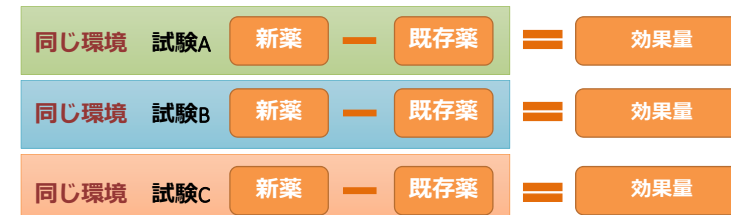
臨床試験単位で取り扱う必要がある

丹後俊郎：メタアナリシス(第2版)，朝倉書店，2016。

メタアナリシスにおけるキーワード：効果量 (effect size)

Simpsonのパラドックスの例：同じ対象、同じ治療法であっても、それぞれの群の患者特性、医師群の特性、診療環境が異なる。

各試験の群間での違い(例：新薬と既存薬での有効率の違い)は、異なった研究においても、同じであると考えるのが自然である。



そのため、メタアナリシスでは、各試験について、それぞれの群の治療効果ではなく、治療効果の違いを取り扱うことが多い。この治療効果の差のことを、効果量(effect size)という。

効果量の種類

相対尺度	
2値	オッズ比 (odds ratio)
	リスク比 (risk ratio)
生存時間	ハザード比 (hazard ratio)
絶対尺度	
2値	リスク差 (risk difference)
連続	平均値の差(mean difference)
	標準化された平均値の差(standardized mean difference)*1

*1：標準化された平均値の差は次式で計算できる。

標準化された平均値の差 = (平均値の差) / (各群の標準偏差の平均値)

なお、Cohen's dとも呼ばれる。

Cohen's dは、標本サイズが小さい場合に過大評価になる可能性があるため、自由度で調整したHedges' gもある。

効果量のグラフィカル表現：Forest plot

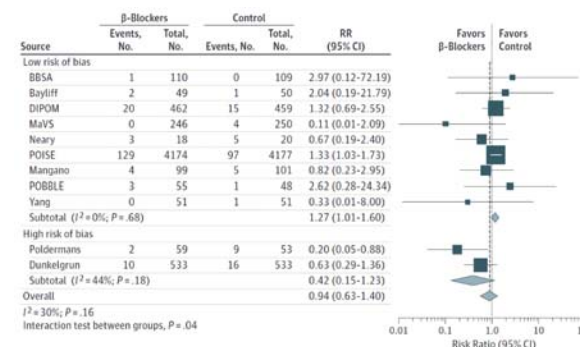
非心臓手術前にβ遮断薬を開始した無作為化比較試験を統合したメタアナリシスの結果である。

Patient：心血管以外の手術を受ける心血管イベントのリスクが高い患者に対して、

Intervention：β遮断薬の投与が、

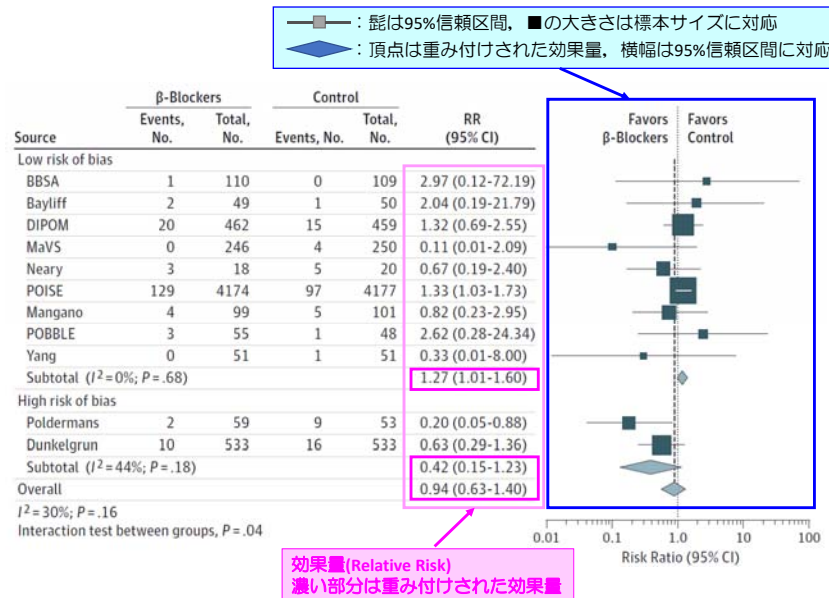
Control：β遮断薬を投与しない患者に比べて

Outcome：術後 30 日あるいは退院時全死亡率(副次的評価項目：非致死的心筋梗塞、脳卒中、低血圧)のリスクを低下させる。



• 個々の研究の効果量及び 95%信頼区間
• 個々の群の標本サイズ
• 重み付けした効果量及び 95%信頼区間
をグラフとともに表したものが**Forest plot**である。

Forest plotのグラフ説明



代表的な統計的モデル

■ 固定効果モデル (Fixed-effect model)

$$\text{各研究の効果量} = \text{共通の効果量} + (\text{偶然}) \text{誤差}$$

効果量は、統合する研究に抛らず同一であると仮定したモデルであり、各試験による効果の違いは(偶然)誤差のみが原因であると考えられる。

■ 変量効果モデル (Random-effect model)

$$\text{各研究の効果量} = \text{共通の効果量} + \text{各研究の偏り} + (\text{偶然}) \text{誤差}$$

誤差の部分

各試験による効果の違いを(偶然)誤差だけでなく、試験によって生じるバラツキにも原因があると考えられる。

各研究の偏りは、**研究の違いによるバラツキ**を持つ。これを**異質性(heterogeneity)**という。

統合解析における効果量の推定

■ 固定効果

連続値(平均値の差)：個々の試験の効果量の標準誤差を用いて重み付けを行う

2値(オッズ比, リスク比, リスク差)：Mantel-Haenszel法, Peto法などがある。

Peto法は、RCTの結果に基づく統合解析以外に用いるべきでない。

異質性がほとんどない場合には、統合された効果量の信頼区間が小さくなる。一方で、異質性が著しい場合には、固定効果モデルの利用は推奨されない。

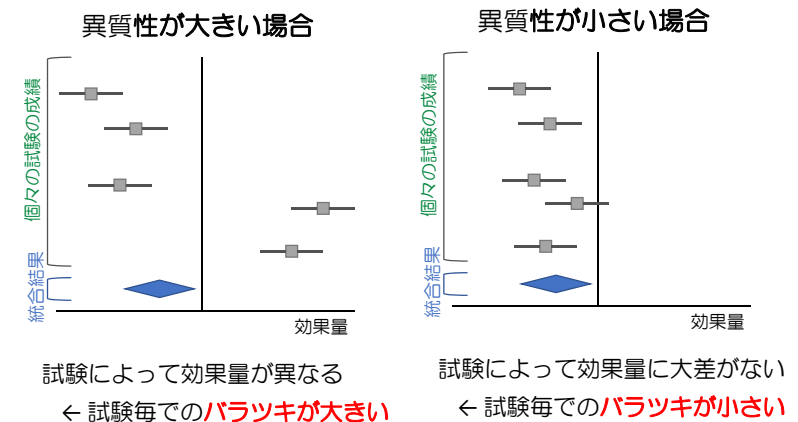
■ 変量効果

DerSimonian-Laird法が一般的に用いられるが、最近では、Bayes流接近法も増えている。

異質性が崩れている場合でも、統合解析が可能。ただし、信頼区間は固定効果モデルよりも広くなる。

異質性を見てしているか否かを精査したうえで、固定効果を用いるか変量効果を用いるかを選択しなければならない。

メタアナリシスの異質性の検討



■ フォレストプロットを視覚的に省察する

個々の試験の効果量の類似性、信頼区間の重なり具合

■ 統計的方法の利用

CochranのQ検定、異質性尺度 I^2 の利用

異質性の評価

■ CochranのQ検定

$$Q = \left[(\text{重み}) \times \{ (\text{試験での効果量}) - (\text{統合した効果量}) \}^2 \right] \text{の合計}$$

Qを利用して検定する。有意な場合には、異質性があると判断し、0.1未満で疑義を持つ必要がある。

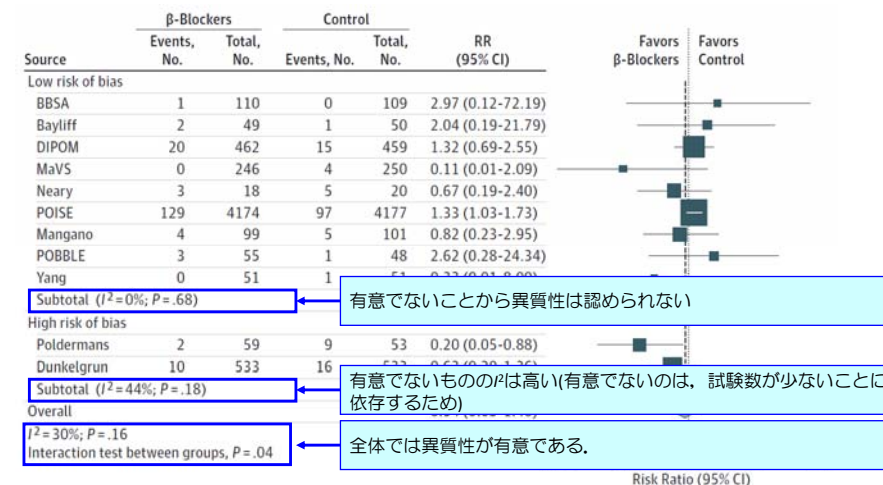
CochranのQ検定は試験数に左右されるため注意が必要(試験数が増えるほど有意になりやすい傾向にある)。

■ 異質性尺度 I^2 の利用

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

0(%)から100(%)までの範囲をとる。Qに対して、試験数で調整を行うため、試験数に左右されない。

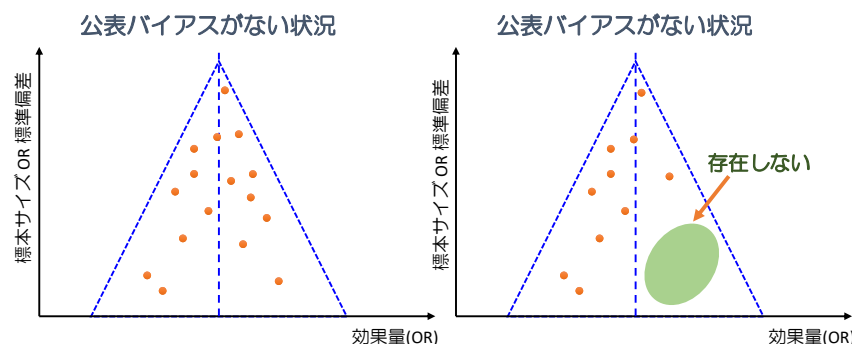
6遮断薬の例示



異質性が認められた場合には、サブセットに分けることで原因を探索することが考えられ、異質な試験のみで評価することが求められる。

公表(出版)バイアスの検討：funnel plot

臨床試験では、結果が有意でない(negative study)のときに論文誌に掲載されない傾向がある(とくに少数例の試験の場合には生じやすい)。このような状況下でのメタアナリシスでは、統合結果にバイアス(偏り)を生じさせる恐れがある。**公表バイアスを評価するためのグラフがfunnel plotである。**



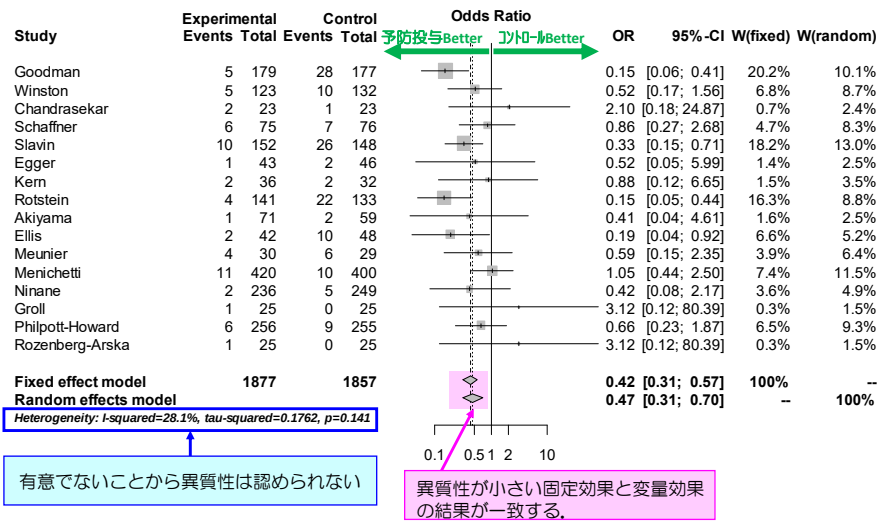
すなわち、funnel plotにおいて対称な場合には、公表バイアスがないと判断され、非対称な場合には、公表バイアスがあると判断される。

メタアナリシスの例示：2値データの場合

抗がん剤治療中の抗真菌剤フルコナゾールの予防投与が全身性真菌症発症の予防に繋がるかどうかのメタアナリシスのデータ(神田,2015)

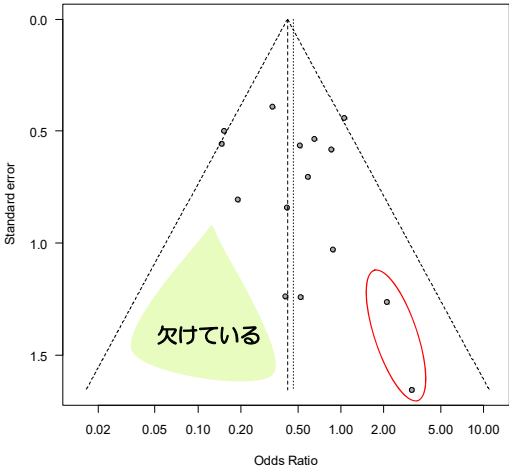
研究名	投与群			非投与群		
	陽性	陰性	例数	陽性	陰性	例数
Goodman	5	174	179	28	149	177
Winston	5	118	123	10	122	132
Chandrasekar	2	21	23	1	22	23
Schaffner	6	69	75	7	69	76
Slavin	10	142	152	26	122	148
Egger	1	42	43	2	44	46
Kern	2	34	36	2	30	32
Rotstein	4	137	141	22	111	133
Akiyama	1	70	71	2	57	59
Ellis	2	40	42	10	38	48
Meunier	4	26	30	6	23	29
Menichetti	11	409	420	10	390	400
Ninane	2	234	236	5	244	249
Groll	1	24	25	0	25	25
Philpott-Howard	6	250	256	9	246	255
Rozenberg-Arska	1	24	25	0	25	25

Forest plot



いずれのモデルでも統合結果(菱形)が1.0を下回っており、かつ、95%信頼区間が1をまたいでいないことから、抗真菌剤フルコナゾールの予防投与が全身性真菌症発症の予防に繋がることが認められる。

Funnel plot



左下が欠けているため、非対称でない傾向を示しているが、ネガティブスタディの小規模な臨床試験が2試験(Groll, Rozenberg-Arska)存在する。これらの試験の95%信頼区間は広く、信頼性が認められない。これらは、統合解析上重みが小さいため、結果にはほとんど影響せず、仮に外した場合には、対称に近くなるかもしれない。

Thank you for your kind attention



メタアナリシスのためのソフトウェア

ソフト名	URL	入手	備考
Comprehensive Meta-Analysis	https://www.meta-analysis.com/	有料	
DistillerSR	https://www.evidencepartners.com/	有料	
JBİ SUMARI	https://www.jbisumari.org/	有料	
Meta-DiSc	http://www.hrc.es/investigacion/metadisc_en.htm	無料	検査値に対するメタアナリシス
MetaXL	https://www.epigear.com/index_files/metaxl.html	無料	Excelで動作するプラグイン
MIX	https://www.meta-analysis-made-easy.com/	有料	Excelで動作するプラグイン
OpenMetaAnalyst	https://www.brown.edu/academics/public-health/research/evidence-synthesis-in-health/open-meta-analyst-tool	無料	
ReviewManeager(RevMan)	http://community.cochrane.org/tools/review-production-tools/revman-5	無料	Cochrane Communityが提供
SAS	https://www.sas.com/ja_jp/home.html	有料	統計解析ソフト
Stata	https://www.lightstone.co.jp/stata/	有料	統計解析ソフト
R	https://www.r-project.org/	無料	統計解析ソフト
EZR	https://www.r-project.org/	無料	Rのプラグイン

システマティックレビュー及びメタアナリシスのチェックリスト：PRISMA

タイトル		
タイトル	1	その報告がシステマティック・レビューなのか、メタアナリシスなのか、あるいはその両方なのかを特定すること。
抄録		
構造化抄録	2	背景、目的、データの情報源、研究の適格基準や参加者や介入、研究の評価および統合方法、結果、限界、結論ならびに主要結果の意味、システマティック・レビュー登録番号などの情報を適宜含んだ、構造化された要約を提供すること。
はじめに		
論拠	3	既知の事項と照らし合わせてレビューの理論的根拠を説明すること。
目的	4	参加者、介入、比較対照、アウトカム、研究デザイン(study design)と関連付けて(PICOS)、取り扱う疑問に関する明確なステートメントを提供すること。
方法		
研究計画書と登録	5	レビューの研究計画書の有無や、そのアクセス可能性とアクセス可能な場所（例：ウェブアドレス）を示し、また入手可能であれば登録番号を含む登録情報を提供すること。
適格基準	6	適格基準として採用された研究特性（例：PICOS, 追跡期間の長さ）や報告特性（例：検討した年数、言語、出版状況）について、理論的根拠を示しながら明示すること。

システマティックレビュー及びメタアナリシスのチェックリスト：PRISMA

情報源	7	検索における全情報源（例：データベースと対象期間、追加的な研究を特定するための著者への連絡）ならびに最終検索日を示すこと。
検索	8	少なくとも1つのデータベースの電子検索式について、使用されたあらゆる“limits”を含め、再現できるくらいに詳細に示すこと。
研究の選択	9	研究の選択過程（すなわち、スクリーニング、適格性、システマティック・レビューへの組み入れ、また、該当する場合はメタアナリシスへの組み入れ）を提示すること
データの抽出過程	10	報告からのデータ抽出方法（例：見本用書式、独立して抽出、2重に抽出）、ならびに研究者からデータを入手し、確認するためのあらゆるプロセスについて説明すること。
データ項目	11	取得されたすべてのデータ項目（例：PICOS, 資金提供者）、ならびにあらゆる仮定や単純化を列挙、定義すること。
個々の研究のrisk of bias	12	個々の研究のrisk of biasを評価するために用いられた方法（これが研究レベルで行われたのか、アウトカムレベルで行われたかの明示を含む）、そしてこの情報があらゆるデータ統合においてどのように使用されるのかを説明すること。
要約指標	13	主な要約指標（例：リスク比、平均差）を提示すること。
結果の統合	14	データの取り扱い方法、そして実施されていれば各メタアナリシスにおける一貫性（例：I ² ）の指標も含め、研究結果の統合方法について説明すること。

システマティックレビュー及びメタアナリシスのチェックリスト：PRISMA

研究全般に関するバイアスのリスク	15	累積エビデンスに影響するかもしれないあらゆるバイアスのリスク評価（例：出版バイアス、研究内での選択的報告）について明示すること。
追加的な分析	16	追加的な分析（例：感度分析またはサブグループ解析、メタ回帰分析）が実施されていれば、その方法を説明し、そのうちのいずれが事前に規定されていたのかを示すこと。
結果		
研究の選択	17	スクリーニングされた研究、適格性が評価された研究、レビューに加えられた研究の件数を示し、各段階での除外の理由について、理想的にはフローチャートを用いて述べること。
研究の特性	18	各研究について、データ抽出が行われる手がかりとなった特性（例：研究の規模、PICOS, 追跡期間）を示し、引用を提示すること。
研究内のrisk of bias	19	各研究のrisk of biasに関するデータ、そして入手可能であれば、アウトカムレベルのあらゆる評価を提示すること（項目12参照）
個々の研究の結果	20	検討対象となったすべてのアウトカム（利益や害）について、研究別に(a) 各介入群に関する簡単な要約データ、(b) 効果推定値と信頼区間を、できればフォレストプロットを付けて提示すること。
結果の統合	21	実施された各メタアナリシスの結果を信頼区間や一貫性の指標を含めて提示すること。
研究全般に関するバイアスのリスク	22	研究全般に関するあらゆるバイアスのリスクの評価の結果を提示すること（項目15参照）。

システマティックレビュー及びメタアナリシスのチェックリスト：PRISMA

追加的な分析	23	追加的な分析〔例：感度分析またはサブグループ解析、メタ回帰分析（項目16参照）〕が実施されていれば、その結果を示すこと。
考察		
エビデンスの要約	24	各主要アウトカムに関して、エビデンスの強さを含め、主な結果について要約すること。またそれらが主要な集団（例：医療提供者、利用者、政策決定者）とどう関係しているか検討すること。
限界	25	研究レベルおよびアウトカムレベルにおける限界（例：バイアスのリスク）、およびレビューレベルにおける限界（例：特定された研究が完全に検索されていない、報告バイアス）について議論すること。
結論	26	結果の一般的解釈を他のエビデンスと関連付けて提示し、今後の研究への影響を示すこと。
資金		
資金	27	システマティック・レビューの資金提供者、ならびにその他の支援（例：データの提供）、そしてシステマティック・レビューにおける資金提供者の役割について説明すること。