

システマティックレビューの
批判的吟味のチェックポイントを
基礎から理解する

京都大学医学部附属病院
精神科神経科
田近 亜蘭

一般的に、エビデンスレベルは このように言われます

(ただし“治療の効果”というテーマについて)

システマティックレビュー/ランダム化比較試験のメタアナリシス

1つ以上のランダム化比較試験

非ランダム化比較試験

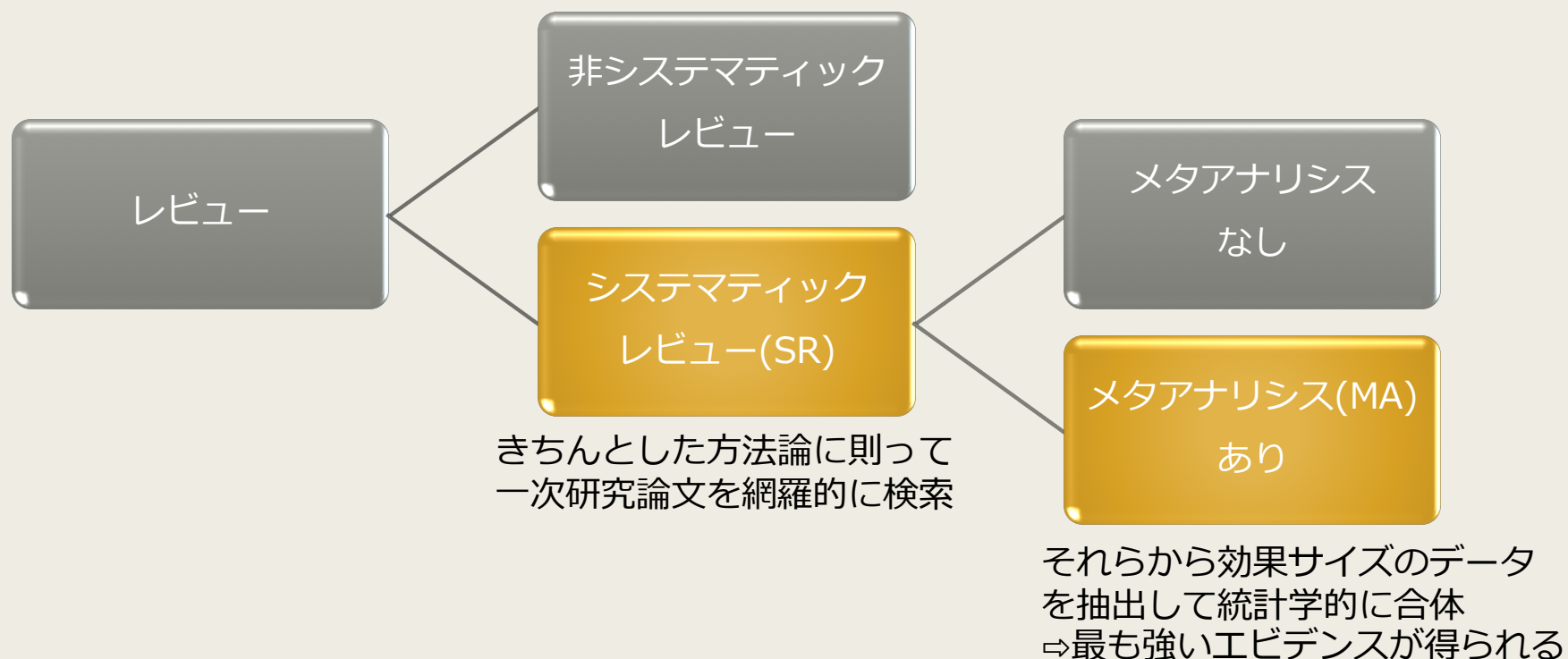
コホート研究

ケース・コントロール研究/横断研究

症例報告/ケース・シリーズ

専門家の意見

レビューの分類



※MAと一口に言っても通常のMA（ペアワイズMA）だけではなく、最近ネットワークMAが増えてきています。しかし今日はネットワークMAについては触れません

※観察研究のMAもありますが、それについても触れません

さて、自分のPECOについて RCTを検索する場合

あなたのPECO $\hat{=}$ RCTの著者のPECO



数多ある論文の中から、できるだけ自分のPECOに近いRCTを検索する
(ピッタリのPECOの論文がないことはよく経験するが、RCTのPECOは比較的是っきりしているので検索しやすい)

では、自分のPECOについて SRを検索する場合

あなたのPECO

≡

SRの著者のPECO



≡ 各RCTの著者のPECO



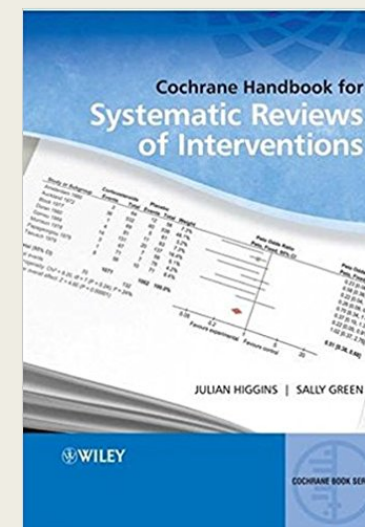
一定の包含基準で、許容されるRCTを集めてくるので、PECOに幅が生じる

できるだけ自分のPECOに近いSRを検索するが、そもそもSRのPECOはRCTと比べて“モヤモヤ”している

SRの批判的吟味をやってみましょう

これは、日々の論文チェックに有意義なだけでなく、ガイドライン作成の際には必須となる知識です

過去の研究会で批判的吟味をしたSRを用いて説明していきます



Khanna. Aripiprazole versus other atypical antipsychotics for schizophrenia, Cochrane Database Syst Rev. 2014

Wijkstra. Pharmacological treatment for psychotic depression, Cochrane Database Syst Rev. 2013

Cipriani 2010 Olanzapine in the long-term treatment of bipolar disorder: A systematic review and meta-analysis, J Psychopharmacol. 2010

京都大学大学院医学研究科 健康増進・行動学分野ホームページ

http://ebmh.med.kyoto-u.ac.jp/index.html



EBM Toolbox



様々なワークシートがダウンロードできます



SRの批判的吟味シート (日本語版) をダウンロード

色々な計算シート

①内的妥当性（比較の妥当性）をチェックし、読み進める価値のある論文であれば

②効果の大きさをNNT (Number Needed to Treat) ※で表現して

③論文の結果を自身の症例に当てはめて考える

系統的レビューの批判的吟味

内的妥当性

(以下の項目について、○、△、×で評価し、その理由を書いてください)

系統的レビューの質（系統的レビューの方法論は妥当か）

1. PECOに定式化された臨床疑問を設定しているか

評価 : ○ △ ×

コメント:

2. 臨床研究を検索する方法が網羅的であったか

評価 : ○ △ ×

コメント:

3. 臨床研究の選択と評価には再現性があったか

評価 : ○ △ ×

コメント:

4. 結果は臨床的に応用できる形で提示されているか

評価 : ○ △ ×

コメント:

得られた効果サイズの質（効果サイズのGRADE 評価）

5. 包含された研究におけるバイアスのリスクは小さいか

評価 : ○ △ ×

コメント:

6. 包含された研究の結果は均質か

結果

- ① Good eventはどのように定義したか。
- ② 治療効果の大きさはNNTにしてどの程度か。
- ③ NNTの95%信頼区間はどうなるか。

単純なITT分析

	Number	Good event	イベント率
介入群 A			
介入群 B			

Worst case scenario

	Number	Good event
介入群 A		
介入群 B		

95%

	下限	上限
OR=		
RR=		
NNT=		

95%

	下限	上限
OR=		
RR=		
NNT=		

(まずは脱落者を non-responder とみなした単純なITT分析を行う。有意差がある場合 scenario も行う)

外的妥当性

(以下の項目について、○、△、×で評価し、その理由を書いてください)

1. あなたの患者は臨床試験の対象となった患者とあまりに異なり、その結果が適用できないほどか

評価 : ○ △ ×

コメント:

2. 臨床試験の結果をあなたの患者に適用すると、どの程度の治療効果が期待されるか

評価 : ○ △ ×

コメント:

3. 代替エンドポイントではなく、臨床的に重要なアウトカムか。副作用は費用は

評価 : ○ △ ×

コメント:

4. あなたの患者の価値観について十分な情報を得ているか。それがこの治療方針で満たされるか

評価 : ○ △ ×

コメント:

自由記載欄:

さて、私の患者さんにはどのようにするのがよいだろうか?

※ NNT : 1つの悪いアウトカムが生じるのを防ぐために治療する必要のある人数のこと
値が小さいほど治療効果は大きいことを意味する

- SRの質（SRの方法論は妥当か）に関して4項目
- 得られた効果サイズの質（効果サイズのGRADE評価）に関して5項目

Grading of Recommendations Assessment, Development, and Evaluation



McMaster大学の Gordon Guyattが提唱

系統的レビューの批判的吟味

内的妥当性
(以下の項目について、○、△、×で評価し、その理由を書いてください)

系統的レビューの質（系統的レビューの方法論は妥当か）		
1. PECO に定式化された臨床疑問を設定しているか		
評価	: ○ △ ×	
コメント:		
2. 臨床研究を検索する方法が網羅的であったか		
評価	: ○ △ ×	
コメント:		
3. 臨床研究の選択と評価には再現性があったか		
評価	: ○ △ ×	
コメント:		
4. 結果は臨床的に応用できる形で提示されているか		
評価	: ○ △ ×	
コメント:		
得られた効果サイズの質（効果サイズのGRADE評価）		
5. 包含された研究におけるバイアスのリスクは小さいか		
評価	: ○ △ ×	
コメント:		
6. 包含された研究の結果は均質か		

1/4

評価	: ○ △ ×	
コメント:		
7. 得られた効果サイズは精確か		
評価	: ○ △ ×	
コメント:		
8. 得られた結果は臨床疑問に設定した患者に当てはまるか		
評価	: ○ △ ×	
コメント:		
9. 包含された研究に報告バイアスは疑われるか		
評価	: ○ △ ×	
コメント:		

自由記載欄:

それぞれの項目について

○ : できている
× : できていない
△ : 記載がない

のいずれかで評価

2/4

【SRの質（SRの方法論は妥当か）】

1. PECOに定式化された臨床疑問を設定しているか

- SRの著者が、漠然とした臨床疑問でなく、PECOに定式化された臨床疑問を設定し、それに基づいてRCTを検索しているか
- ObjectiveやIntroductionの最後に記載されていることが多い



2. 臨床研究を検索する方法が網羅的であったか

【出版された論文について】

- デジタルデータベース

それぞれの分野で重要なデータベースの検索が欠けていないか



- ハンドサーチ

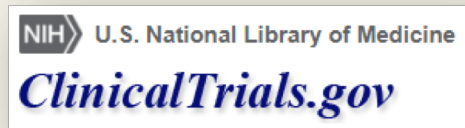
先行研究のReferenceを検索して、漏れているものがないか

- 英語以外の言語の論文はどこまで検索する？

【出版されていないデータについて】

- Registry

登録されて出版されていないRCTがないか，主な登録サイトを検索



- 学会発表の抄録
- 学位論文
- 製薬会社に眠っているデータ
- 個人的コンタクト
- その他…



これら全部を検索することはなかなかできないが，頑張れば頑張るほど，SRの価値は上がる



あなたは未出版データを持っているはずだ！

3. 臨床研究の選択と評価には再現性があったか

- 包含される研究のスクリーニング, データ抽出, バイアスの評価など, 質的な判断が必要な際に, 複数の人でチェックしているか

例.

Two review authors independently reviewed all identified publications...

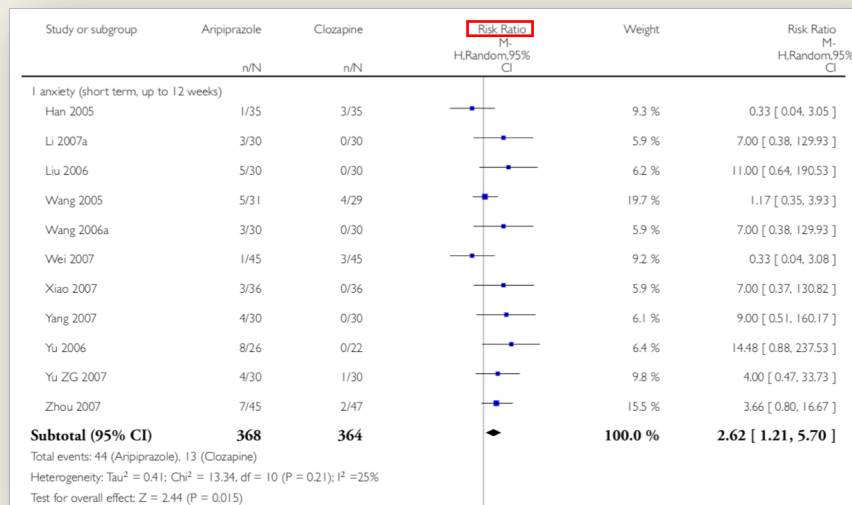
Any disagreement was resolved by consensus discussion with another review author.

こういった, お約束の文言が記載されているはず

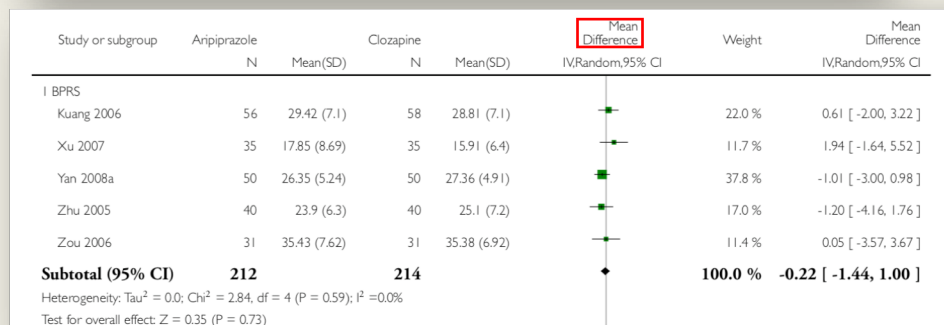


4. 結果は臨床的に応用できる形で提示されているか

- 二値変数であればリスク比やオッズ比, 連続変数であればMD (Mean Difference) やSMD (Standardized Mean Difference) など, 効果サイズの指標で結果が提示されているか



例. レスポンダーの割合の場合
⇒リスク比



例. 評価尺度の点数の場合
⇒Mean Difference

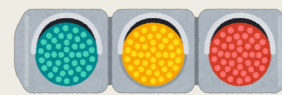
【得られた効果サイズの質（効果サイズのGRADE評価）】

5. 包含された研究におけるバイアスのリスクは小さいか

- SRに含まれる個々のRCTのバイアスのリスクを評価し、それが小さいか

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding (performance bias and detection bias): of participants	Blinding (performance bias and detection bias): of personnel	Blinding (performance bias and detection bias): of outcome assessors	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other bias
Anton 1990	?	?	+	+	?	+	+	+
Bruijn 1996	?	?	+	+	+	+	?	+
Meyers 2009	+	?	+	+	+	+	+	?
Mulsant 2001	?	?	+	+	+	+	?	+
Rothschild 2004a	?	?	+	+	?	?	?	+
Rothschild 2004b	?	?	+	+	?	?	?	+
Spiker 1985	?	+	+	+	+	+	?	?
Spiker 1988	+	?	+	+	?	?	?	+
van den Broek 2004	+	?	+	+	+	+	?	+
Wijkstra 2010	+	+	+	+	+	+	+	?
Zanardi 1996	?	?	?	?	?	+	?	?
Zanardi 2000	+	?	?	?	+	+	?	?

- Random sequence generation (selection bias)
割り付けは無作為に行われていたか
- Allocation concealment (selection bias)
割り付けは隠蔽化されていたか
- Blinding (performance bias and detection bias): of participants
患者はブラインドであったか
- Blinding (performance bias and detection bias): of personnel
治療者はブラインドであったか
- Blinding (performance bias and detection bias): of outcome assessors
評価者はブラインドであったか
- Incomplete outcome data (attrition bias)
アウトカム評価ができていないケースが多くなかったか
- Selective reporting (reporting bias) (=結果に影響するほどの脱落がなかったか)
報告すべきアウトカムは全て報告されているか
- Other bias (=都合のいいものだけを報告していないか)
その他のバイアスがなかったか



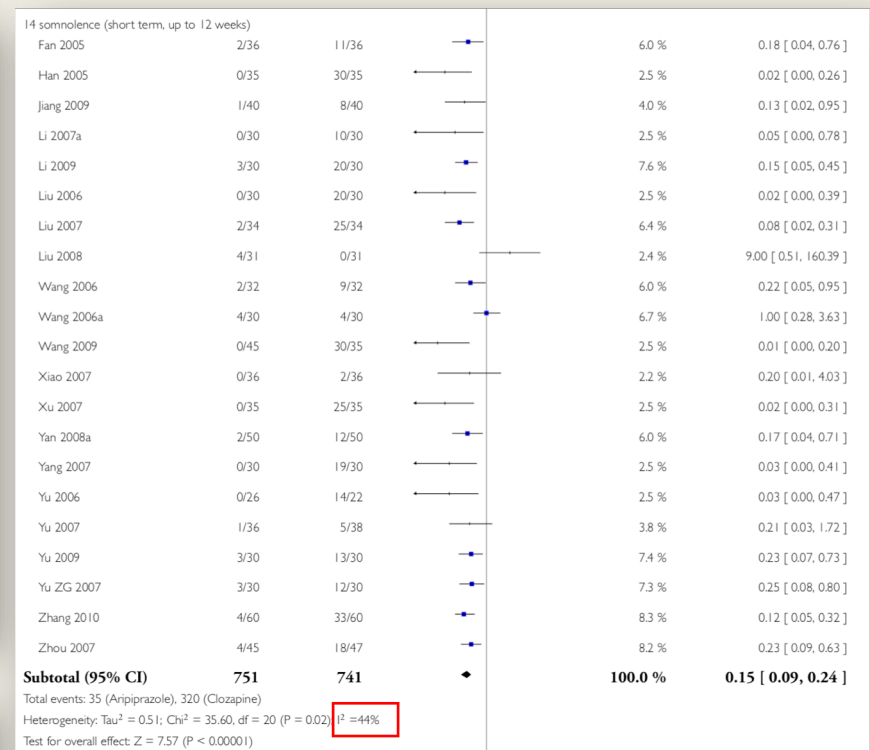
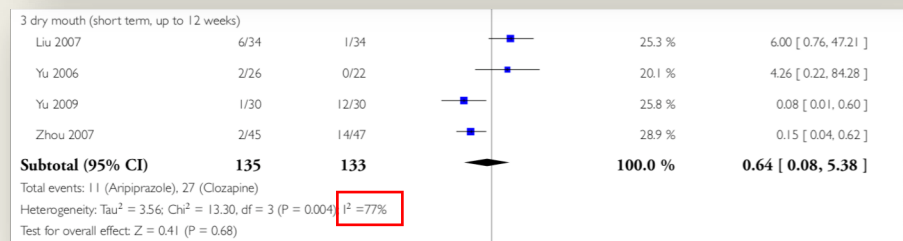
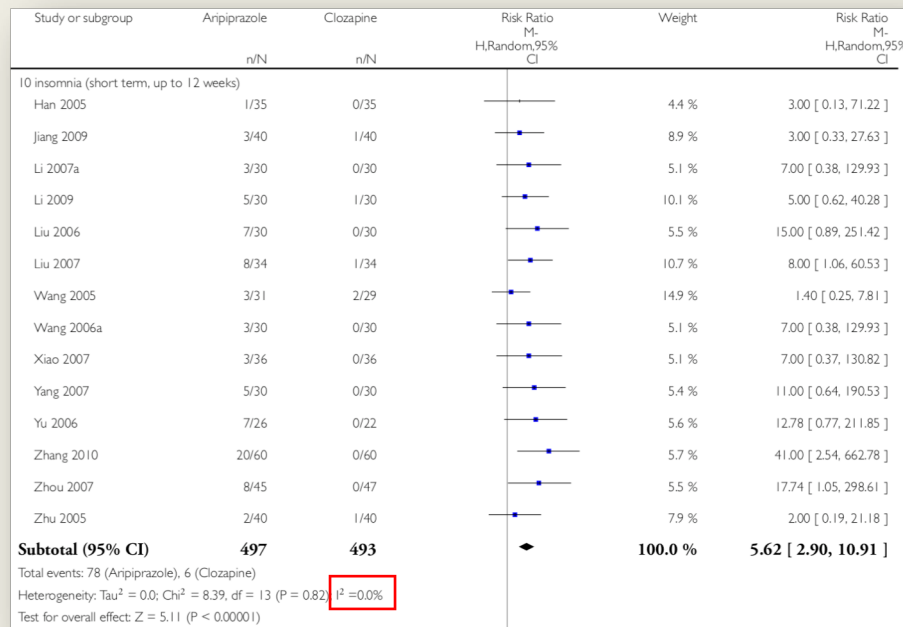
Low risk of bias
Unclear risk of bias
High risk of bias

赤信号が多過ぎやしませんか？



6. 包含された研究の結果は均質か

■ 含まれるRCTの結果の異質性を，統計学的・視覚的にチェック



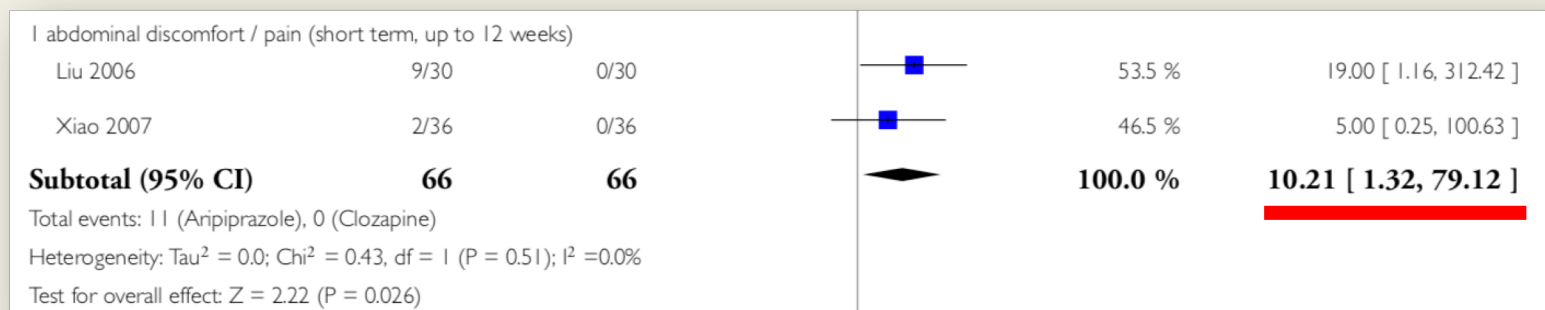
■ 統計学的異質性：I²

I ²	解釈
0～40%	異質性は問題ないかも
30～60%	中等度の異質性があるかも
50～90%	かなりの異質性があるかも
75～100%	重大な異質性

■ 視覚的異質性：各RCTの効果サイズの点推定値が逆を向いていないか

7. 得られた効果サイズは精確か

- サンプル数やイベント数が少なく95%信頼区間の幅が広い場合、効果サイズが不精確になってしまう



8. 得られた結果は臨床疑問に設定した患者にあてはまるか

- SRの著者が設定したPECOと、実際に集まった各RCTのPECOとの相違が大きくないか

この部分のはなし

あなたのPECO



≡

SRの著者のPECO



≡ 各RCTの著者のPECO

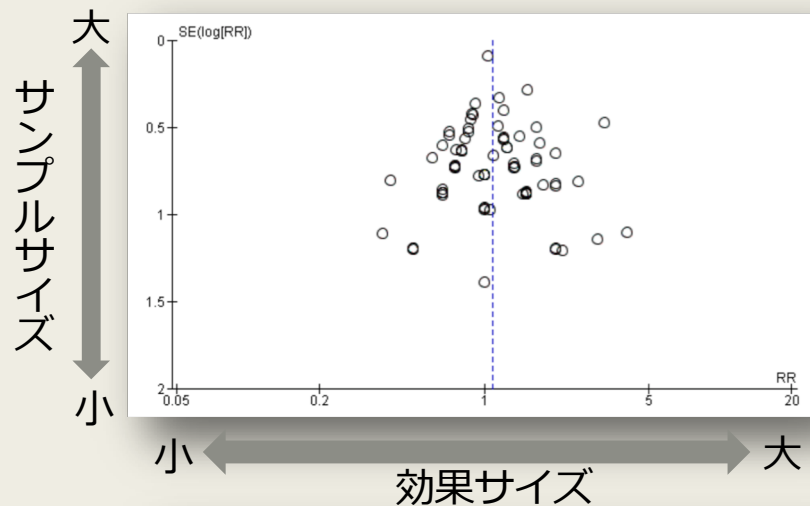


SRでは、そもそも自分のPECOに合致したRCTを探してくるので、大きくずれることはないはず

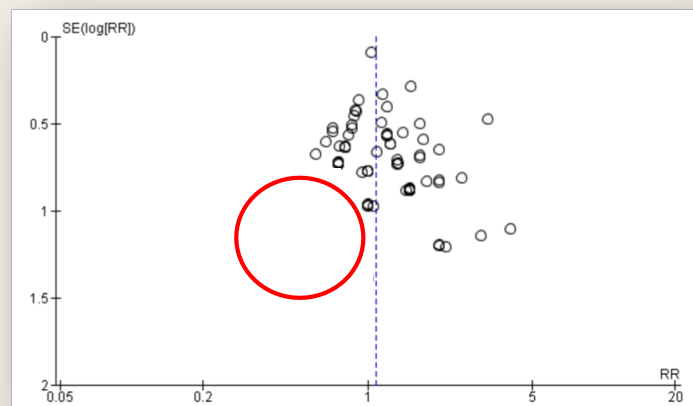
しかし、ガイドライン作成の際には、当初設定したPECOと、見つかったRCTのPECOがずれることは多い
(直接性が下がる)

9. 包含された研究に報告バイアスは疑われるか

- Funnel plot analysis (漏斗図分析) などで出版バイアスがチェックされているか



サンプルサイズの小さい研究ほど効果サイズがばらつくため、左右対称に漏斗型に分布するはず



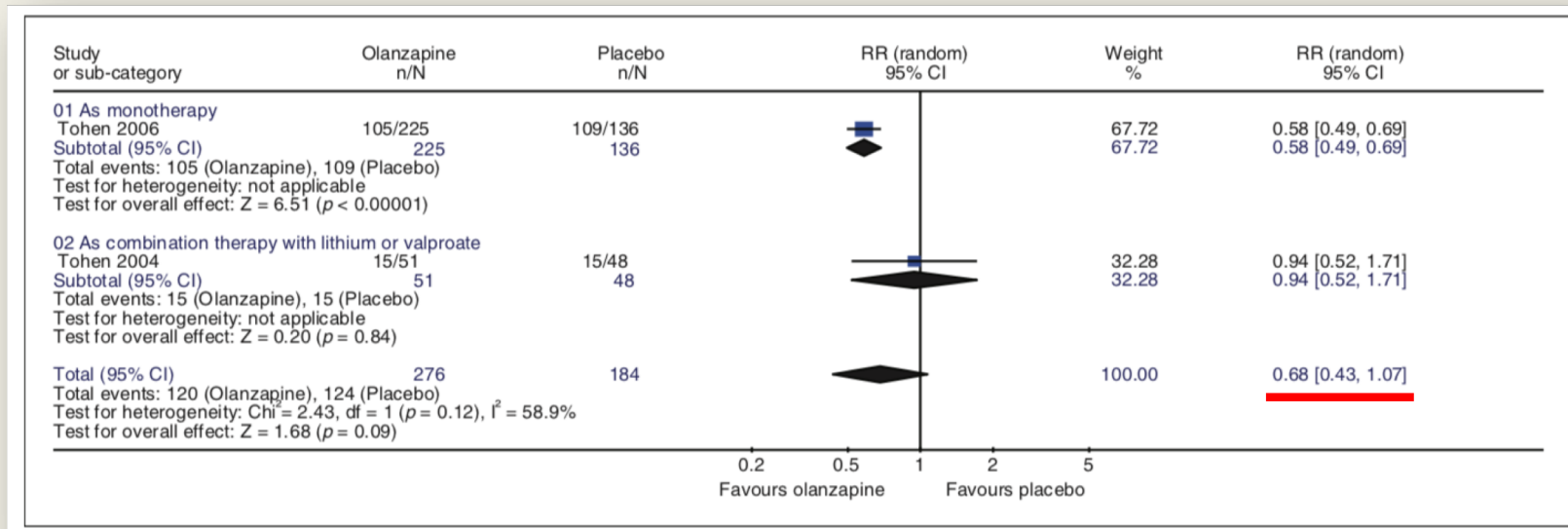
謎の空白エリアがあれば、本来あるべきはずの研究が欠けている？

(参考) GRADE評価

G. Guyatt et al. / Journal of Clinical Epidemiology 64 (2011) 383–394

Study Design	Quality of Evidence	Lower if	Higher if
Randomized trial →	High	Risk of bias 項目5. -1 Serious バイアスのリスク -2 Very serious	Large effect +1 Large +2 Very large
	Moderate	Inconsistency 項目6. -1 Serious 異質性 -2 Very serious	Dose response +1 Evidence of a gradient
Observational study →	Low	Indirectness 項目8. -1 Serious 直接性 -2 Very serious	All plausible confounding +1 Would reduce a demonstrated effect or
	Very low	Imprecision 項目7. -1 Serious 精確性 -2 Very serious	+1 Would suggest a spurious effect when results show no effect
		Publication bias 項目9. -1 Likely 出版バイアス -2 Very likely	

RCTであれば、QualityはHighからスタートし、これらの項目に該当していればランクダウンしていく



例えば、こういった結果からNNTを計算したい場合

RR = 0.68 (←メタアナリシスで得られたRRを使用)

CER (Control Event Rate) = (109+15)/(136+48)=124/184=0.67
(↑便宜上、各研究の単純合計人数を使用)

EER (Experimental Event Rate) = CER×RR=0.67×0.68=0.46

NNT (Number Needed to Treat) = 1/(CER-EER)=1/(0.67-0.46)=4.76

NNTの95%信頼区間についてはRRの下限0.43、上限1.07について同様に計算する

外的妥当性

(以下の項目について、○、△、×で評価し、その理由を書いてください)

1. あなたの患者は臨床試験の対象となった患者とあまりに異なり、その結果が適用できないほどか

評価 : ○ △ ×

コメント:

2. 臨床試験の結果をあなたの患者に適用すると、どの程度の治療効果が期待されるか

評価 : ○ △ ×

コメント:

3. 代替エンドポイントではなく、臨床的に重要なアウトカムか。副作用は費用は

評価 : ○ △ ×

コメント:

4. あなたの患者の価値観について十分な情報を得ているか。それがこの治療方針で満たされるか

評価 : ○ △ ×

コメント:

自由記載欄:

さて、私の患者さんにはどのようにするのがよいだろうか?

外的妥当性は主治医にしかわからない!

自分の患者さんを思い浮かべながら、存分に想像力をはたらかせる

それぞれの項目について

- : あてはまる
- × : あてはまらない
- △ : 判断できない

のいずれかで評価

1. あなたの患者は臨床試験の対象となった患者とあまりに異なり、その結果が適用できないほどか

- もしあなたの患者さんが、SRに含まれるRCTのエントリー現場にいたとしたら、この研究に組み入れられていたか
- もしこのSRの包含基準を満たしていなかったとしてもその違いは臨床的に許容される範囲か

例. SRの包含基準が18歳から70歳までだった場合、あなたの患者が72歳だったら結果はあてはまらない？

2. 臨床試験の結果をあなたの患者に適用すると、どの程度の治療効果が期待されるか

- この論文の対象患者と比べて、あなたの患者は、よりアウトカムを発生しやすいか

例. あなたの患者は薬剤反応性が悪いとすると、論文の結果を当てはめる際には、はなし半分くらいに思っておかないといけないかも
(= NNTは論文の値よりも大きくなるイメージ)

逆に薬剤反応性が良いのであれば、NNTは小さくなるイメージ

3. 代替エンドポイントではなく、臨床的に重要なアウトカムか。副作用は 費用は

- 代替エンドポイント：本当に知りたい重要なエンドポイントの身代わりのエンドポイント

例. 骨折の発生 vs. 骨密度（代替）

心筋梗塞の発症 vs. コレステロール値（代替）

- 代替エンドポイントかどうかを判別するには「もしそのアウトカム以外の何も変わらないとした場合、あなたはその治療を受けたいと思うか」を考えればいい
- 益と害の兼ね合いで考える
- どんなに有効な治療でも、日本で適応がなかったり、費用がかかり過ぎたりする場合は？？

4. あなたの患者の価値観について十分な情報を得ているか. それがこの治療方針で満たされるか

- どんなに抗うつ薬の有効性がSRで証明されたとしても、「私、絶対に抗うつ薬は飲みたくないんです！」と言っている患者さんに対してその結果は適用できない
- それとも、その論文の結果を示して患者さんを説得する？

これら外的妥当性のポイント1~4を踏まえて、
「私の患者さんにはどのようにするのがよいだろうか？」
を総括する

さあ，この研究会で 批判的吟味の練習を してみませんか？



今後の予定は以下のとおりです（原則，第1木曜日 ですが，時々週がずれます）

	5月	7月	9月	11月	1月	3月
前半	レクチャー (RCT)	症例報告	症例報告	レクチャー (RCT)	症例報告	症例報告
後半	批判的 吟味	レクチャー (文献検索 演習)	批判的 吟味	批判的 吟味	レクチャー (メタ)	批判的 吟味

+
忘年会🍷

毎回，メールでご案内をお送りします
ご参加，お待ちしております