

# ガイドラインの エビデンス評価

埼玉医科大学 形成外科・美容外科

時岡一幸

# ガイドライン作成作業（抜粋）

1. CQの設定
2. エビデンスの収集
  - データベース検索、文献のスクリーニング
3. エビデンスの評価（個々の文献）
  - 二次スクリーニングで残った採用文献を個々に評価する
  - 評価シート（SR-5、SR-6）の作成
4. エビデンスの統合（エビデンス総体の評価）
  - 個々の文献評価をまとめて、エビデンスの総合評価を行う
  - 評価シート・エビデンス総体（SR-7）の作成

## 3 & 4. エビデンスの評価と統合

- CQに合わせて収集した個々の文献を評価したのち、CQ全体としてエビデンスを統合する
- システマティックレビューと同様の作業

### 1. 定性的システマテックレビュー ⇒ 全てのCQに対して行う

- 各研究のデータを質的に統合する



可能なら

### 2. 定量的システマテックレビュー（メタアナリシス MA）

- 各研究のデータを統計学的に統合する
  - 症例数がまとまった文献が複数ある場合に行うことができる
- 
- 研究デザインによって（介入研究・観察研究）、評価項目および評価シートが異なる

## 個々の文献評価を行ったのち、その結果を統合する

個々の文献の評価



エビデンスの統合

個別研究		バイアスリスク*						上昇要因**		非直接性*				リスク人数（アウトカム率）								効果指標（種類）		効果指標（値）		信頼区間	
		選択バイアス	実行バイアス	検出バイアス	症例減少バイアス	その他																					
研究コード	研究デザイン	背景因子の差	ケアの差	不適切なアウトカム測定	不完全なフォローアップ	交絡の調整	その他のバイアス	まとめ	量反応関係	効果減弱交絡	効果の大きさ	まとめ	対象	介入	対照	アウトカム	まとめ	対照群分母	対照群分子	(%)	介入群分母	介入群分子	(%)				
Czaja 2008	コホート研究	-1	-1	0	-1	-2		-2	0	0	0	0	-2	-2	-1	0	-2	8	7	87.5	72	33	45.8	RR	0.52	0.16-0.89	
Czaja 2008	コホート研究	-2	-1	0	-1	-2		-2	0	0	0	0	-2	-2		0	-2	NA	NA	NA	206	76	36.9	NA	NA	NA	

定量的SRでは必須

エビデンス総体								リスク人数（アウトカム率）											
アウトカム	研究数 研究デザイン	バイ* アスリスク	非一* 貫性	不精* 確性	非直* 接性	その他* (出版バイアスなど)	上昇要因* (観察研究)	対照群分母	対照群分子	(%)	介入群分母	介入群分子	(%)	効果指標（種類）	効果指標統合値	95 % 信頼区間	エビ* デンスの強さ	重要* 性	コメント
生化学的・組織学的寛解	コホート研究 ／3	-2	0	-1	-2			8	7	0.9	540	101	0.187	NA	NA	NA	弱(C)	8	PSL + AZA または PSL 単独は寛解導入に有効と考えられる。

エビデンス総体は「CQ設定」に記載されたアウトカムごとにまとめる

# 個々の文献の評価

バイアスリスク

非直接性

上昇要因

リスク人数

# 個々の文献評価（介入研究用）

著者名  
発行年

ポップアップリスト  
より選択

①バイアスリスク：研究のデザイン、実行、分析における偏り

②非直接性：CQと文献の差

⑤リスク人数・効果指標  
文献の症例数など

個別研究		バイアスリスク*										非直接性*					リスク人数（アウトカム率）							（種類）	* 効果指標（値）	信頼区間			
		選択バイアス		実行バイアス	検出バイアス	症例減少バイアス	その他																						
研究コード	研究デザイン	ランダム化	コンシールメント	盲検化	盲検化	ITT	不完全報告	アウトカム報告	選択的アウトカム報告	中止	早期試験中	その他のバイアス	まとめ	対象	介入	対照	アウトカム	まとめ	対照群分母	対照群分子	(%)	介入群分母	介入群分子	(%)					
Slayer 2021	RCT	0	0	-1	-1	-1	0	0	0	0	0	-1	-1	-1	0	-1	0	-1	0	-1	1520	407	26.8	1515	408	26.9	RR	1.006	0.89-1.13
													▲						▲										
コメント（該当するセルに記入）																													
Slayer 2021				不十分	不十分	未施行						交絡因子疑い		海外の報告		対象は健康者													

④それぞれのまとめを評価する

下の方にあるコメント欄に理由などを表記する

# 個々の文献評価（観察研究用）

著者名  
発行年

ポップアップリスト  
より選択

①**バイアスリスク**：研究のデザイン、実行、分析における偏り

②**非直接性**：CQと文献の差

③**上昇要因**  
エビデンスを上げる要因

⑤**リスク人数・効果指標**  
文献の症例数など

個別研究		バイアスリスク*						上昇要因**				非直接性*				リスク人数（アウトカム率）								（種類） ***効果指標	（値） 効果指標	信頼区間
研究コード	研究デザイン	背景因子の差	ケアの差	不適切なアウトカム測定	不完全なフォローアップ	交絡の調整	その他のバイアス	まとめ	量反応関係	効果減弱交絡	効果の大きさ	まとめ	対象	介入	対照	アウトカム	まとめ	対照群分母	対照群分子	(%)	介入群分母	介入群分子	(%)			
Yuzuriha 2019	症例集積	-2	-2	-1	0	0	0	-2	0	0	0	0	0	0	-1	-1	-1	24	15	62.5	19	14	73.6			
								▲				▲					▲									
コメント（該当するセルに記入）																										
Yuzuriha 2019	後ろ向き比較研究	裂型が異なる	術者が複数	術者が測定											PB法と比較	計測が異なる										

④**それぞれのまとめを評価する**

下の方にあるコメント欄に理由などを表記する

# 評価項目の概要

- 個々の参考文献をGRADEシステムに準拠して、評価する
  - Grading of Recommendations Assessment, Development and Evaluation
- ① バイアスリスク
  - 研究のデザイン、実施、分析における偏りによって、誤った結果や結論を導くリスク
- ② 非直接性
  - CQと文献の間における、PICOの差（症例、介入方法など）
- ③ 上昇要因
  - （観察研究において）エビデンスを上げる要因
- ④ まとめ
  - ①～③の総合評価
- ⑤ リスク人数・効果指標 ⇒ 定量的SRが対象
  - 症例数、発生頻度、平均値など

①②は減点

③は加点する



# 評価項目：研究デザインによる違い

- 研究デザインによって評価項目が異なる
  - 介入研究の評価はより厳密
  - アウトカムに及ぼす影響の程度を考慮して評価する

		バイアスリスク	上昇要因	非直接性
データ総合型研究	SR、MA	エビデンス総体として評価（後述）		
介入研究	RCT	○		○
	準RCT	○		○
観察研究 （疫学）	コホート研究	○	○	○
	症例対照研究	○	○	○
	横断研究		○	○
観察研究	症例集積研究	○*	○	○
	症例報告	（原則的に個別評価は不要）		
その他	総説・専門家の意見	（原則的に個別評価は不要）		

\*比較対照がない研究では不要

# (参考) 研究デザイン

作成ファイル



**介入研究とは** 「人を対象とする医学系研究に関する倫理指針」（2015年施行）より抜粋

- ① 通常の診療を超えた医療行為に関する研究（保険未収載、適応外使用の治療）
- ② 通常の診療と同等の医療行為であっても、被験者の集団を2群以上に分け、それぞれに健康事象に影響を与える要因（治療、診断、予防など）の割り付けを行って、効果等を群間で比較する研究

# ①バイアスリスク（観察研究）

リスク低：0  
中：-1  
高：-2

## ・研究の厳密さ（いい加減さ）の評価

選択バイアス	背景因子の差	比較される群間で対象の背景因子に差があるか？ 例）年齢、性別、重症度、併存疾患、等
実行バイアス	ケアの差	比較される群間で介入は一定か？ 例）治療へのアドヒアランス、他の併用療法、治療時期*、等
検出バイアス	不適切な アウトカム測定	比較される群間でアウトカム測定の差があるか？ 例）測定者が介入に気づく、測定・評価の信頼性が低い、等
症例減少バイアス	不完全な フォローアップ	比較される群間で症例の減少に差があるか？ 例）脱落例、アウトカムデータが不精確、等
その他のバイアス	交絡因子の調整	交絡因子の存在はあるか？ 十分な調整ができていないか？ 例）多変量解析が未施行
	その他	利益相反（営利団体の関与など）

\* 介入の施行時期が異なると、使用器材やlearning curveなどの影響が無視できない

\*\* 交絡因子：要因とアウトカムの両者に影響を及ぼす因子

# バイアスリスク評価の例（観察研究）

アウトカムに及ぼす影響を考慮して判定する

## 1. 選択バイアス

- 口蓋形成術の研究、群間で裂型分布が異なる  $\Rightarrow -1 \sim -2$

## 2. 実行バイアス

- 群間で治療時期が異なる  $\Rightarrow -1 \sim -2$
- 群間で術者の経験年数が異なる  $\Rightarrow -1 \sim -2$

## 3. 検出バイアス

- 測定者が医療提供者  $\Rightarrow -2$ （恣意的な測定の可能性）

# ① バイアスリスク（介入研究）

リスク低	: 0
中	: -1
高	: -2

- **選択バイアス：比較される群間で対象症例の背景因子に偏りがないか？**
  - **ランダム化：無作為割り付け（ランダム化）の有無・質**
    - 乱数表、コンピューター ⇒ 0
    - 誕生日、受診日、カルテ番号 ⇒ -1 ~ -2
    - ベースライン不均衡（症例の背景因子が異なる） ⇒ -1 ~ -2
  - **コンシールメント：割り付けの隠蔽化**
    - （医療現場以外での）中央割り付け ⇒ 0      上記以外 ⇒ -1 ~ -2
- **実行バイアス：比較される群間で介入（医療行為など）の実行に系統的な差があるか？**
  - **盲検化（医療提供者・参加者）**
    - 盲検化なし ⇒ -1 ~ -2
- **検出バイアス：比較される群間でアウトカム測定に系統的な差があるか？**
  - **盲検化（アウトカム測定者）**
    - 盲検化なし ⇒ -1 ~ -2

# ①バイアスリスク（介入研究）

リスク低：0
中：-1
高：-2

- **症例減少バイアス：比較される群間で症例の減少に系統的な差があるか？**

- **ITT**（Intention To Treat 治療企図分析）\*

- あり ⇒ 0      なし ⇒ -1~-2

- **アウトカム不完全報告**：アウトカムが不完全な症例（脱落など）の頻度

- 通院中止や副作用のため脱落した症例が群間で異なる ⇒ -1~-2

## その他

- 選択的アウトカム報告：測定されたアウトカムが全て報告されているか？ ⇒ -1~-2

- 早期試験中止：予定の症例数を満たす前に試験が終了していないか？ ⇒ -1~-2

- その他のバイアス

- 利益相反（営利組織からの資金提供など） ⇒ -1~-2

- 多変量解析によって交絡因子\*\*の調整を行っていない ⇒ -1~-2

- 単施設での臨床試験 ⇒ -1~-2

\*ITT：アウトカムが不完全な症例（脱落など）を除外せず、最初の割り付け通りに解析する手法

\*\*交絡因子：要因とアウトカムの両者に影響を及ぼす因子

例）飲酒者では肺癌の発生率が高いが、そもそも飲酒者の喫煙率が高い

# バイアスリスク評価の例（介入研究）

アウトカムに及ぼす影響を考慮して判定する

## 1. 選択バイアス

- ランダム化の方法・隠蔽化の記載がない ⇒ -2
- 生まれ月で割り付け（奇数月・偶数月） ⇒ -1

## 2. 検出バイアス

- 測定者の盲検化なし ⇒ アウトカムに及ぼす影響の強さによって判定
  - アウトカム：死亡 ⇒ 0（誰が評価しても大差なし）
  - アウトカム：運動機能 ⇒ -1（測定者によって異なる可能性あり）
- 測定者が医療提供者 ⇒ -2（恣意的な測定の可能性）

## ②非直接性 【介入研究・観察研究で共通】

低：0  
中/疑い：-1  
高：-2

- CQと参考文献の間で、PICOの各要素に差があるか？

<b>対象</b> 研究対象の違い	人種*、年齢、性別、重症度、合併症、など 例) CQと文献で裂型が異なる
<b>介入</b> 介入方法の違い	投与方法・量、術式の詳細、後療法など 例) 術式は同じだが、文献では補助療法が追加されている
<b>対照</b> 比較される対照の違い	比較対照（健常者、異なる裂型・術式など）、間接比較** 例) CQは術式AとBの比較、文献では術式Aと未手術例の比較
<b>アウトカム</b> アウトカム測定の違い	計測方法、評価基準など 例) CQでは再手術の頻度で評価、文献ではCT値で評価

\*海外の文献では総じて非直接性が存在する。アウトカムに及ぼす影響の程度によって判定する

例) 術後の瘢痕は日本人と白色人種で大きく異なる、言語形態の違い、など

\*\*間接比較：介入AとBの直接比較ではなく、共通の対照を介した比較



### ③上昇要因 【観察研究のみ】

非常に大きい：+2  
大きい：+1  
なし：0

- 介入によって生じるアウトカムを増強させる因子の存在  
⇒効果が絶大であり、エビデンスが上昇する要因となりうる

量反応関係	介入や要因（投与量、放射線被ばく量など）が用量依存性の反応を示す 例）ワーファリンの投与量と抗凝固作用の関係
効果減弱交絡	介入や要因の効果を減弱させる交絡因子の存在 例）介入群の方が対照群より症例の重症度が高いが、それでも成績が良い
効果の大きさ	介入の効果が非常に大きい リスク比（RR）：5以上または0.2以下⇒+2 2以上または0.5以下⇒+1

**リスク比**：介入群と対照群におけるアウトカム発生率の比  
例）5以上または0.2以下 ⇒発生率が5倍以上または1/5以下

## ④まとめ

リスク低：0
中：-1
高：-2

- ①～③の各項目の評価を統合して判定する
  - 各項目の-2や-1の割合を参考にする
  - 割合が絶対ではない！（アウトカムに及ぼす影響を考慮して判定する）

（参考）定量的評価の方法 相原守夫「診療ガイドラインのためのGRADEシステム」

総合評価	各項目のバイアスリスク
低（0）	高（-2）がない 中（-1）が3個以下
中（-1）	高（-2）が1個 いずれの項目にも高（-2）はないが、中（-1）が4個以上
高（-2）	上記以外 高（-2）が2個以上

## ⑤ リスク人数・効果指標

- 定量的システマティックレビューの際に必須となる
  - 介入研究（RCTなど）が主な対象
  - 二値変数** ⇒ 症例数を記載      **連続変数** ⇒ 平均値・標準偏差を記載

in the GPP group compared to SBG (risk ratio: 11.51, 95% CI: 3.39-35.15).

二値変数用 リスク人数（アウトカム率）						（種類） 効果指標	（値） 効果指標	信頼 区間
対 照 群 分 母	対 照 群 分 子	(%)	介 入 群 分 母	介 入 群 分 子	(%)			

論文の丸写しでよい

RR	リスク比
OR	オッズ比
RD	リスク差
NA	記載なし

ポップアップリストから選択

# （参考） 効果指標

$$\text{リスク比 (RR)} = \frac{Cl_1}{Cl_0} = \frac{a/m}{b/n}$$

$$\text{リスク差 (RD)} = Cl_1 - Cl_0 = \frac{a}{m} - \frac{b}{n}$$

$$\text{オッズ比 (OR)} = \frac{Cl_1/(1 - Cl_1)}{Cl_0/(1 - Cl_0)} = \frac{ad}{bc}$$

オッズ : ある事象が起きる確率 (p) と起きない確率の比 (1-p)

オッズ比 : 2群間におけるオッズの比

$$\text{相対リスク減少 (RRR)} = 1 - RR$$

非暴露群（対照群）と比較した場合の、暴露群（介入群）における発生率の減少率

$$\text{絶対リスク減少 (ARR)} = -RD$$

非暴露群（対照群）と比較した場合の、暴露群（介入群）における発生率の減少値

アウトカム	介入群：m	対照群：n
あり	a	b
なし	c	d
発生率	$Cl_1 = a/m$	$Cl_0 = b/n$

# エビデンスの統合

バイアスリスク

非直接性

上昇要因

非一貫性

不精確性

出版バイアス

# エビデンス総体

## アウトカム・研究デザインごとに分ける

アウトカムは「CQ設定」ファイルを参照

## ②エビデンスの強さ

## 4 段階 (A~D)

### ③重要性

ファイル「CQ設定」を参照

#### ④リスク人数・効果指標

## 症例数のまとめ等

### ①評価項目

## エビデンス総体

アウトカム	研究デザイン／ 研究数	バイアスリスク*	非一貫性*	不精確性*	非直接性*	その他 (出版バイアスなど)	上昇要因(観察研究)*	対照群分母	対照群分子	(%)	介入群分母	介入群分子	(%)	効果指標 (種類)	効果指標統合値	95 % 信頼区間	エビデンスの強さ**	重要性***	コメント
6 ヶ月後の総死亡	RCT ／3	0	0	0	-1	0		1960	456	0.223	1976	464	0.23	RR	1.0101	0.905- 1.139	A	9	差がない。
6 ヶ月後の依存性 (mRS0-2)	RCT ／3	0	0	0	-1	0		1960	1226	0.626	1976	1176	0.6	RR	0.9231	0.827- 1.041	B	7	mRS0-2 がアウトカム
1 週間以内の脳出血	RCT ／3	0	0	0	-1	0		1961	29.5	0.015	1977	144.5	0.07	RR	4.4371	2.013- 9.777	A	9	脳出血例のほとんどが死亡

コメント（該当するセルに記入）

定性SRの場合、④は不要

[illegible]

# ①評価項目

<b>バイアスリスク</b> risk of bias	個々の研究で評価したバイアスリスクのまとめ ⇒次頁へ
<b>非直接性</b> indirectness	個々の研究で評価した非直接性のまとめ ⇒次頁へ
<b>非一貫性</b> inconsistency	研究間の結果のばらつき、異質性（heterogeneity）
<b>不精確性</b> imprecision	サンプルサイズやイベント数が少ないために、アウトカムの効果判定に疑問が生じる
<b>出版バイアス</b> publication bias	研究が選択的に報告されることによって、根底にある介入の益と害の効果が過小評価または過大評価される
<b>上昇要因</b>	個々の研究で評価した上昇要因のまとめ（観察研究のみ） ⇒次頁へ

個々の評価に  
なかった項目

# ① バイアスリスク・非直接性・上昇要因

- 個々の文献評価における、「まとめ」の統合
  - 個々の文献の－2や－1の割合を参考にする
  - 割合が絶対ではない！（アウトカムに及ぼす影響を考慮して判定する）

（参考）定量的評価の方法 相原守夫「診療ガイドラインのためのGRADEシステム」

エビデンス総体の バイアスリスク	個々の文献のバイアスリスクまとめ
<b>低（0）</b>	全ての研究が低（0）である
<b>中（－1）</b>	全ての研究が低（0）または中（－1）である 1件の高（－2）があるが寄与率*25%以下である
<b>高（－2）</b>	高（－2）の研究が2個以上 高（－2）の研究が1個であるが、かなりの寄与率である

\* 寄与率（重み付け）：SRの中で各研究の情報が占める割合



# ①非一貫性・不精確性・出版バイアス

- 文献ごとの結果のばらつき、サンプルサイズなどに関する評価

<b>非一貫性</b> inconsistency	研究間の結果のばらつき 結果の点推定値（平均値、中間値など）、信頼区間が異なる
<b>不精確性</b> imprecision	サンプルサイズやイベント数が少ないために、アウトカムの効果判定に疑問が生じる 症例数が最適情報量（optimal information size : OIS）を満たすか？ 無効果の効果推定値95%信頼区間に含まれるか？
<b>出版バイアス</b> publication bias	研究が選択的に報告されることによって、根底にある介入の益と害の効果が過小評価または過大評価される 研究のアウトカムに偏りがある 製薬会社などのスポンサー

定性的SRと定量的SRでは評価方法が異なる

⇒定量的SRでは専用ソフトを用いて計算する

# 定量的SR（メタアナリシス MA）

- 複数の研究の症例数、点推定値などを統計学的に統合する

- 専用のソフトを用いる

**R program**

<https://cran.r-project.org/>



CRAN  
Mirrors  
What's new?  
Task Views  
Search

About R

The Comprehensive R Archive Network

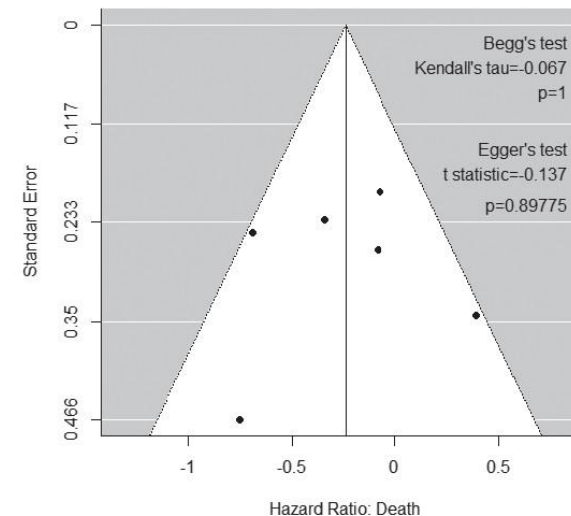
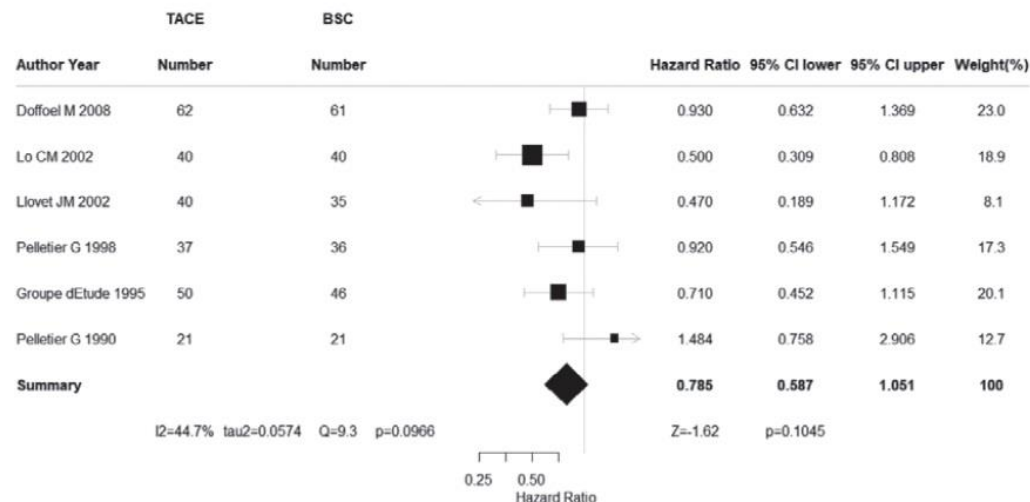
Download and Install R

Precompiled binary distributions of the base system and contributed packages, Windows and Mac users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

- SR論文で見られる以下のような図表が作成できる

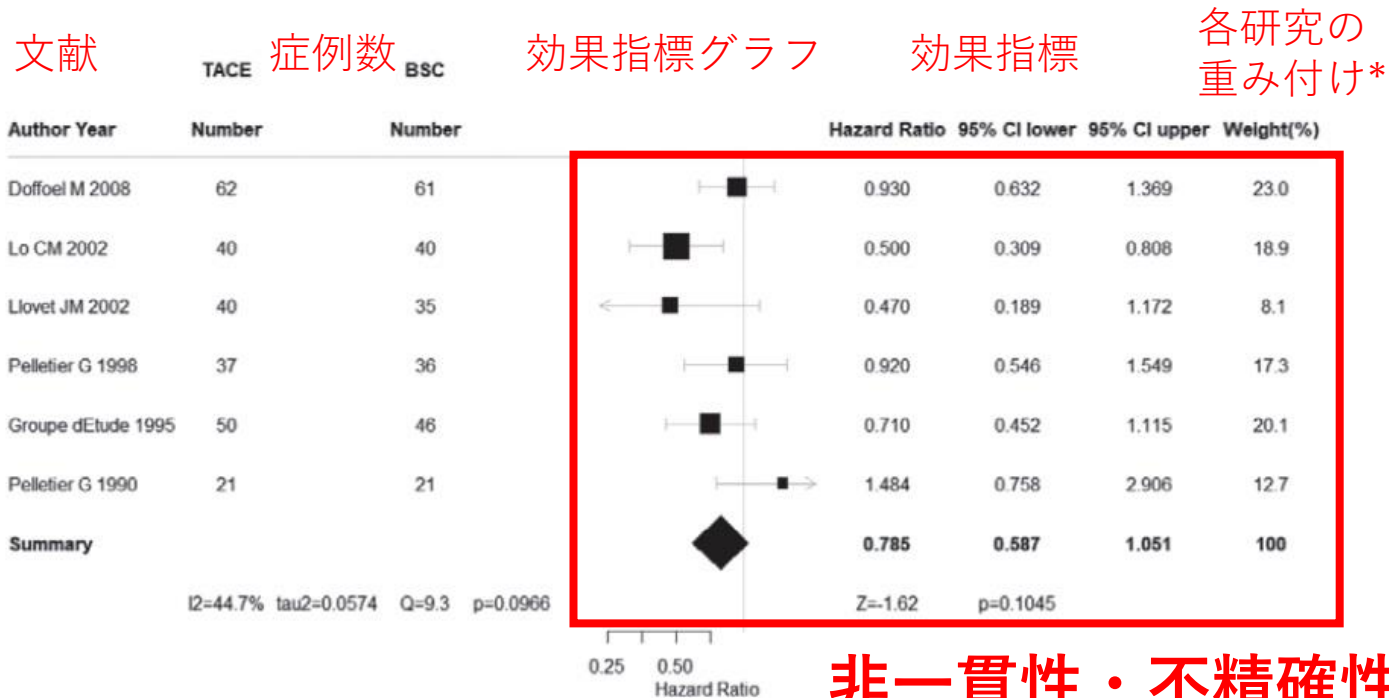


「Minds マニュアル 2020」より

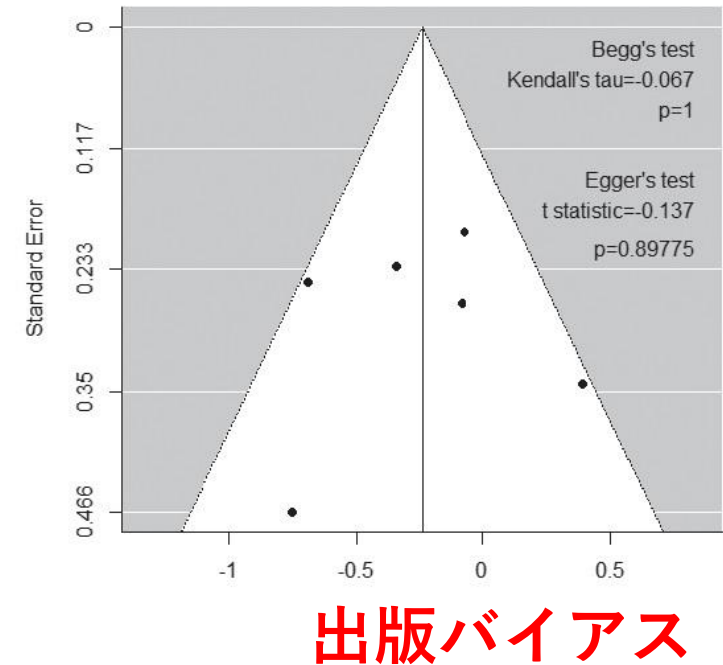
# ①非一貫性・不精確性・出版バイアス

- ・ ソフトで作成した図表を見て判定する（SR論文で出てくるやつ）

## フォレストプロット



## ファンネルプロット



\* 個々の研究の情報量（症例数など）に依存する  
グラフの■の面積として反映される

「Minds マニュアル 2020」より

# 非一貫性 (inconsistency)

- 定義：結果に関する説明のつかない異質性 (heterogeneity)
  - 研究間で治療効果の推定値が大きく異なる
- 原因
  - 対象                      例) 症例の背景リスクが高い群の方が薬物効果が高い
  - 介入                      例) 薬剤の用量が多いほど効果が大きい
  - アウトカム              例) 追跡機関が長いほど治療効果が減少する
  - 研究方法                例) バイアスリスクが高い研究と低い研究で効果が異なる
- 評価基準

1) 点推定値が研究間で異なり，その相違がかなり大きい	目視
2) 各信頼区間の重なりが，ほとんどまたは全くない	目視
3) $I^2$ 値（研究間の差異による点推定値のばらつきの割合）が大きい	統計的手法
4) 異質性検定（コクランQ検定）において， $p < 0.05$ で帰無仮説を棄却できる	統計的手法

# 非一貫性 (inconsistency)

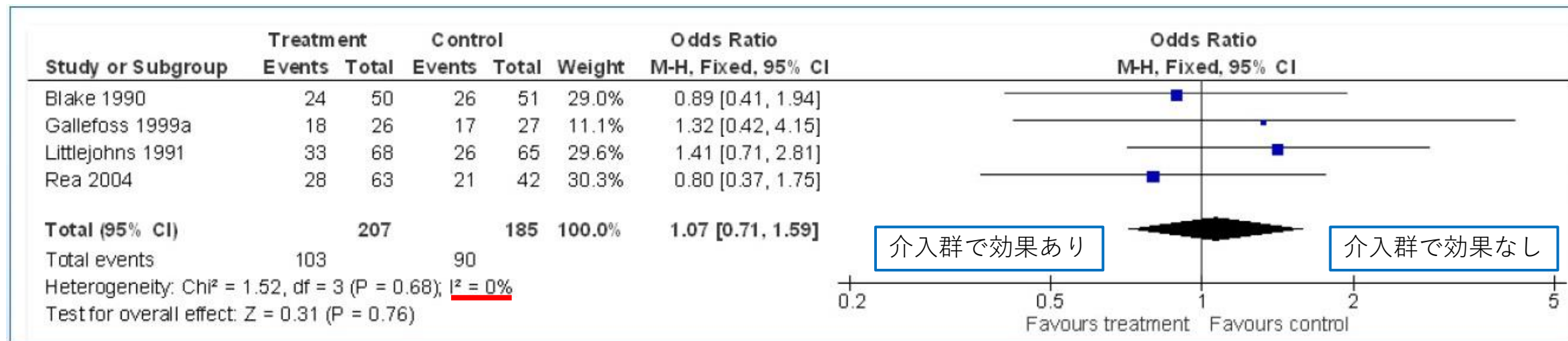
- ・ フォレストプロットで判定する

個々の文献

症例数  
【介入】 【対照】

効果指標  
オッズ比 [95%信頼区間]

指標効果グラフ



$I^2$ 値で判定

< 40%	小さい
30～60%	中等度
50～90%	かなり大きい
75～100%	非常に大きい

各研究の点推定値・95%信頼区間のばらつきを目視で評価

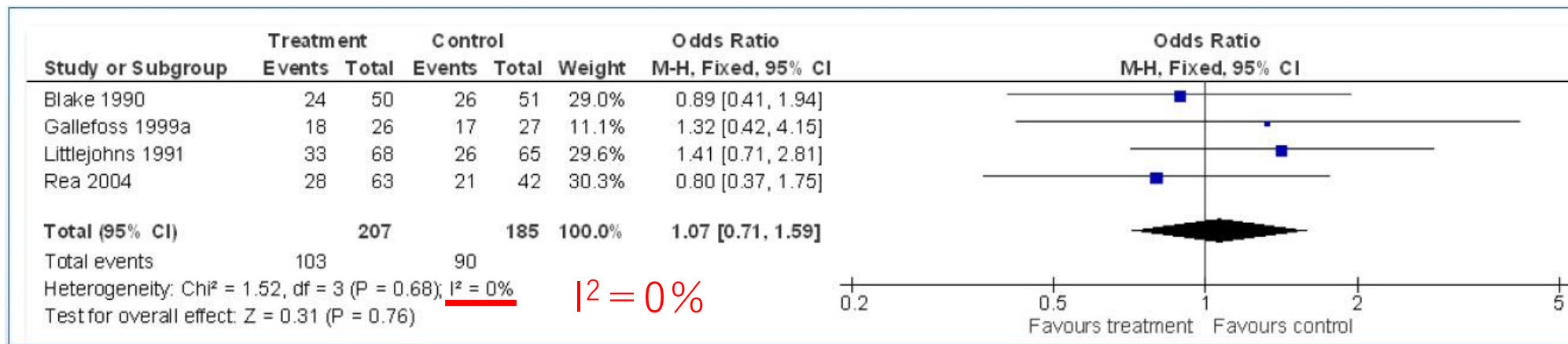
「Minds エビデンス総体の確実性」より

# 非一貫性の例

「Minds エビデンス総体の確実性」より

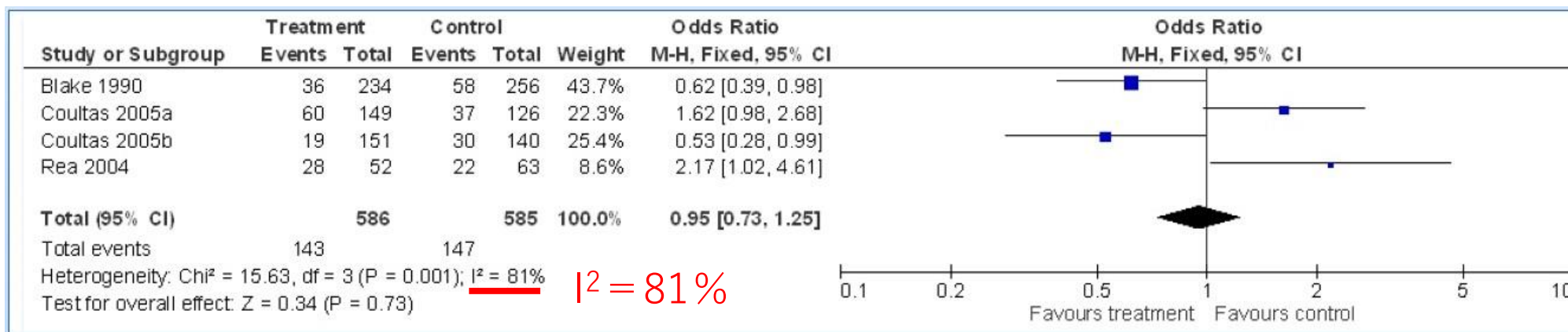
例1) 点推定値のばらつきはあるが、信頼区間はそれなりに重なっている

判定



非一貫性なし  
(0)

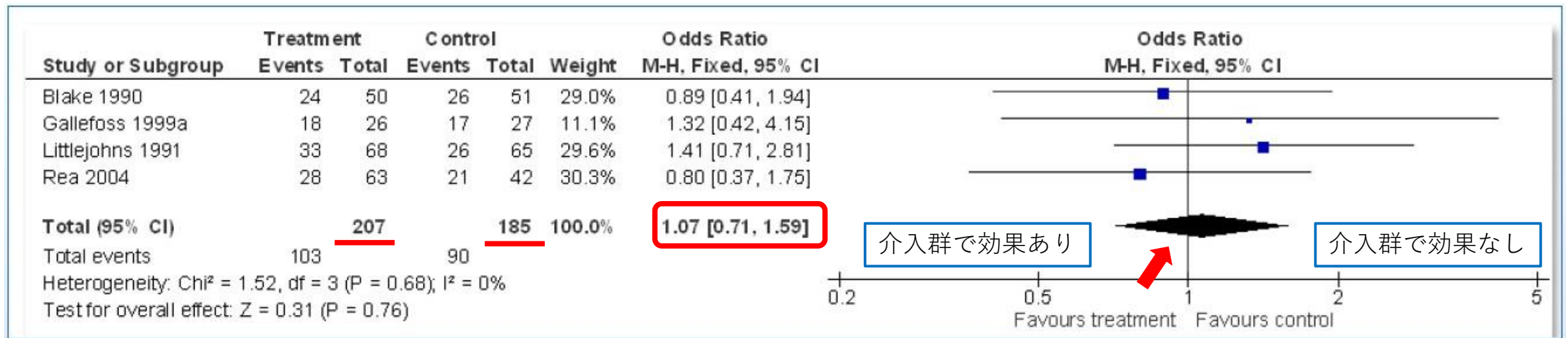
例1) 点推定値のばらつきが大きく、信頼区間の重なりがほとんどない



深刻な非一貫性  
(-1~-2)

# 不精確性 (Imprecision)

- (SR全体の) 症例数などが少ないために、効果推定値の信頼区間が広い
- 判定基準
  - ① 症例数は最適情報量 (optimal information size : OIS) を満たすか？
  - ② 無効果の効果推定値が統合された95%信頼区間に含まれるか？



①専用のツールでOISを計算する

②信頼区間の統合値が中央線をまたぐか

「Minds エビデンス総体の確実性」より



# 不精確性 (Imprecision)

## ① 症例数は最適情報量 (optimal information size : OIS) を満たすか？

- サンプル数計算ソフトを用いる <http://stat.ubc.ca/~rollin/stats/ssize/b2.html#42>

- ☒ Calculate Sample Size (for specified Power)
- ☐ Calculate Power (for specified Sample Size)

Enter a value for p1:

Enter a value for p2:

- ☐ 1 Sided Test
- ☒ 2 Sided Test

Enter a value for  $\alpha$  (default is .05):

Enter a value for desired power (default is .80):

The sample size (for each sample separately) is:

例) 2種類の口蓋形成術による瘻孔発生率の比較  
これまでの発生率は20%程度だが、新たな介入による発生率の減少を検討する研究

p1 (対照群発生率) = 0.20

p2 (介入群発生率) =  $p1 \times 0.75^* = 0.15$

defaultのままでよい

**OISが計算される**

\*  $0.75 = 1 - 0.25$  (RRR: 相対リスク減少) として設定

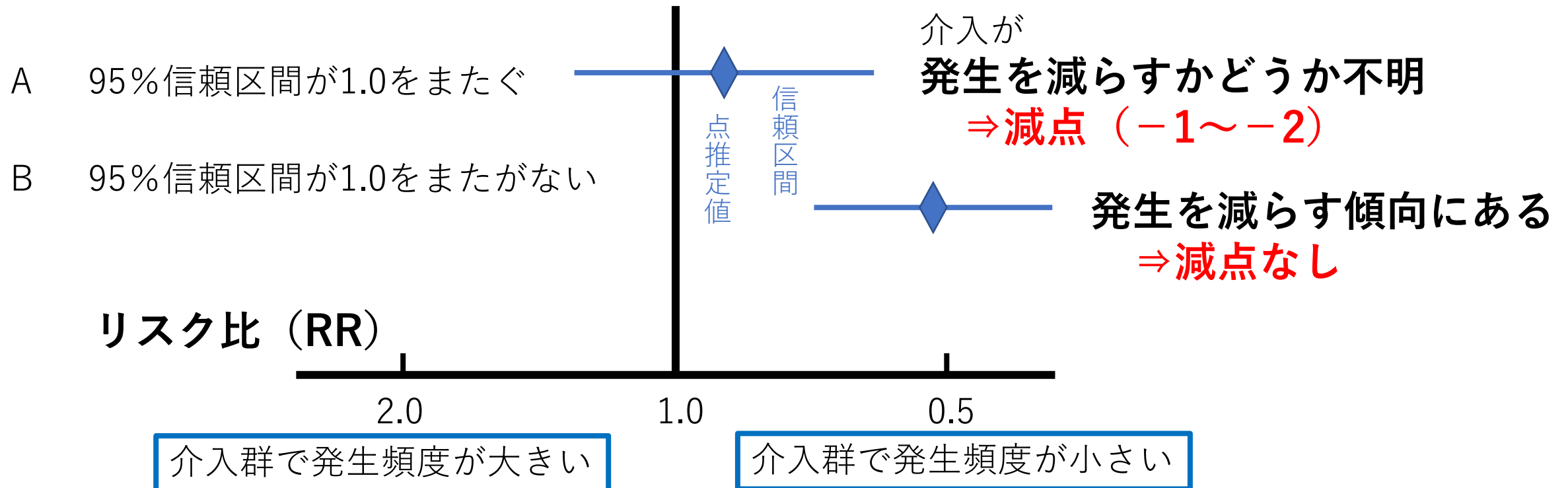
介入によって得られる現実的な効果として、20 ~ 30% (通常25%) のRRRを設定する  
この例では「介入群では瘻孔発生率が25%減少する」の意味



# 不精確性：信頼区間の判定

② 無効果の効果推定値（リスク比＝1、リスク差＝0）が統合された95％信頼区間に含まれるか

例）介入群と対照群の口蓋瘻孔発生率の比較

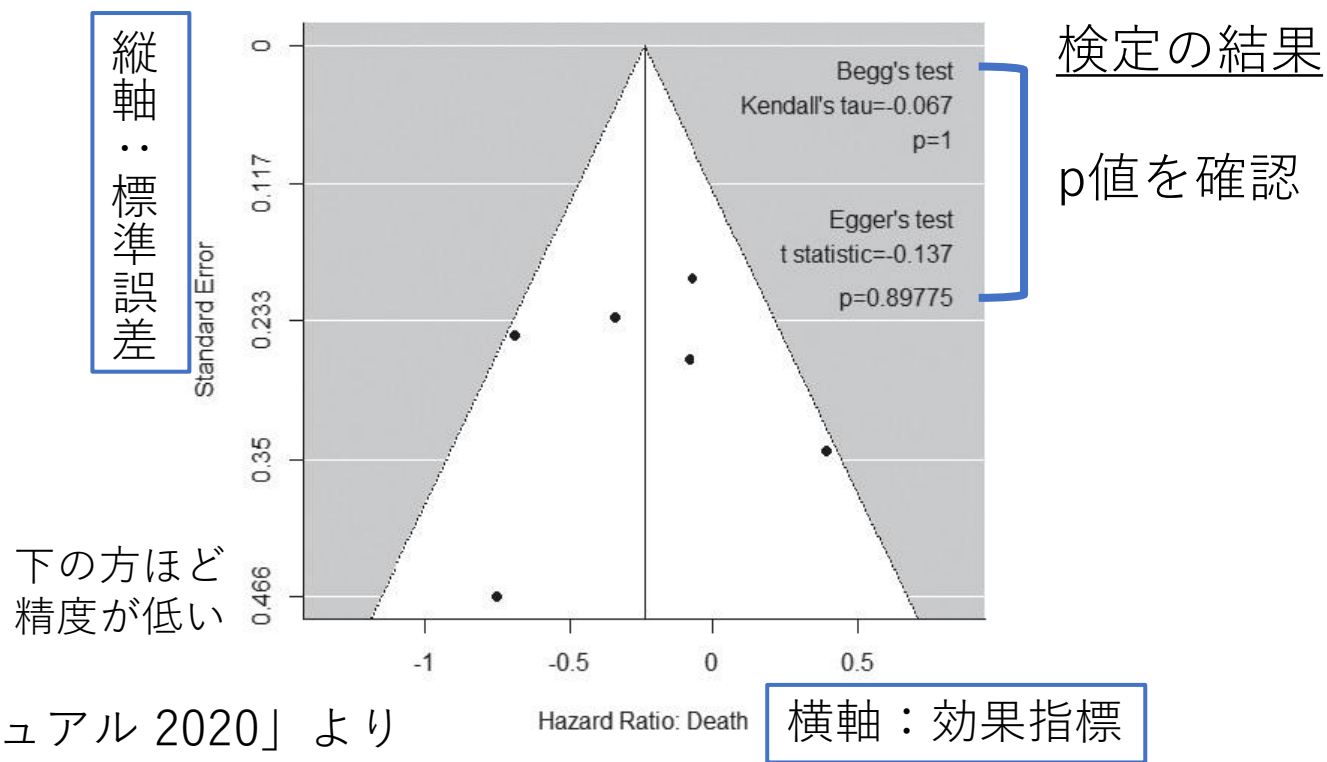


# 出版バイアス (publication bias)

- 研究が選択的に報告されることによって、根底にある介入の益と害の効果が過小評価または過大評価される
  - 例) スポンサー等の存在

- 評価方法：**ファンネルプロット**

- 左右対称性の目視確認
  - 各点が個々の研究を示す
- 統計学的検定 (Egger、Beggなど)



# ①非一貫性・不精確性・出版バイアス

- 定性的SRの評価方法

<b>非一貫性</b> inconsistency	研究間の結果のばらつき 例) 同じ術式でも、再手術の頻度が大きく異なる 例) 計測値の平均値・中間値が大きく異なる
<b>不精確性</b> imprecision	サンプルサイズやイベント数が少ないために、アウトカムの効果判定に疑問が生じる 例) 症例数の少ない症例集積報告のみ
<b>出版バイアス</b> publication bias	研究が選択的に報告されることによって、根底にある介入の益と害の効果が過小評価または過大評価される 例) 負のアウトカムに関するデータがない (少ない) 例) 製薬会社などのスポンサーがある

## ②エビデンスの強さ

- エビデンス総体の評価項目を総合して評価する

- |                 |                                |
|-----------------|--------------------------------|
| <b>強 (A)</b>    | 効果の推定値が推奨を支持する適切さに強く確信がある      |
| <b>中 (B)</b>    | 効果の推定値が推奨を支持する適切さに中程度の確信がある    |
| <b>弱 (C)</b>    | 効果の推定値が推奨を支持する適切さに対する確信は限定的である |
| <b>非常に弱 (D)</b> | 効果の推定値が推奨を支持する適切さにほとんど確信できない   |

- 原則的に強 (A) からスタートして、評価項目に応じてグレードダウンする
- 交絡因子の影響が大きいと考えられる場合は、弱 (C) からスタートする  
例) 術式による治療成績の評価において、裂型 (交絡因子) の調整が行われていない

### ③（アウトカムの）重要性

- 診療上の意思決定によって影響されるアウトカムの重要度
  - 患者から見たアウトカムの重要性を重み付けする
- 「CQ設定」ファイルより転記する

重要性：大								
1	2	3	4	5	6	7	8	9
意思決定に重要性は低い アウトカム			意思決定に重要だが 重大ではないアウトカム			意思決定に 重大なアウトカム		

例) 口蓋形成術

入院  
期間

手術  
回数

顎  
発育

言語  
成績

口  
蓋  
瘻  
孔

気  
道  
閉  
塞

# 個々の文献の評価：付記

SR・MA文献の評価方法

# システマティックレビュー（SR）が抽出された場合

- SRやメタアナリシス（MA）はそれ自体が複数の研究を統合した論文であるため、扱いが異なる
- PICOが一致、かつ、2023年の発行
  - ① 統合結果をそのまま利用する
    - エビデンス総体のファイルに評価内容を記載する
- PICOが不一致、または、2022年以前の発行
  - ② 文献に採択されている元の研究論文からCQに適合するものを選択してシステマティックレビューを行う



二次スクリーニング後に残った文献に含まれていなければ、  
ハンドサーチとしてCQの採用文献に追加する

# (参考) SRの正式な扱い

## Mindsの推奨

- SRの評価を行う
  - 質           ⇒CQとPICOが合致するか？ 文献検索戦略は適切か？
  - 発行年      ⇒SRの出版後に新たな研究がどれくらいあるか？
- 評価に応じて対応する
  - ① 統合結果をそのまま利用する
  - ② 論文で採択されている元の研究データを再評価し、元の研究からCQ に適合するものを選択してシステマティックレビューを行う
  - ③ 同じ文献検索戦略を用いて得られた文献で新たにシステマティックレビューを行う
  - ④ 新しい研究を追加してメタアナリシスを行うか定性的な統合を行う
  - ⑤ 文献検索戦略の一部を用いて得られた文献で新たにシステマティックレビューを行う

\* ③～⑤は作業が大変なため、日形会の方針を得て、前頁の方針とした



# SRの評価方法

- SRはそれ自体が複数の研究の統合結果であるため、**エビデンス総体**のシートを用いる

エビデンス総体		①評価項目						②定量的SR（MA）の数値						エビデンスの強さ 重要性			
アウトカム	研究デザイン／ 研究数	バイアスリスク	非一貫性	不精確性	非直接性	その他 （出版バイアスなど）	上昇要因（観察研究）	リスク人数（アウトカム率）				効果指標（種類）	効果指標統合値	95％信頼区間	エビデンスの強さ	重要性	コメント
								対照群分母	対照群分子	(%)	介入群分母						
鼻咽腔閉鎖機能	SR/1																
顎発育	SR/1																

論文の中にヒントがあります ⇒次頁へ

論文の中にヒントがあります ⇒次頁へ

- アウトカムと重要性は「CQ設定」をみて記入する
  - CQに含まれないアウトカムは記載しなくてよい
  - 1つのSR論文の中に複数のアウトカムがある場合は個別に記載する

# SRの評価項目

- エビデンス総体の評価項目と同じ

①評価項目	
バイアスリスク	個々の研究で評価したバイアスリスクのまとめ
非直接性	個々の研究で評価した非直接性のまとめ
非一貫性	研究間の結果のばらつき
不精確性	サンプルサイズやイベント数が少ないために、アウトカムの効果判定に疑問が生じる
出版バイアス	研究が選択的に報告されることによって、根底にある介入の益と害の効果が過小評価または過大評価される
②定量的SR（MA）の数値	
リスク人数	SRの症例数（介入群・対照群）、アウトカム率
効果指標など	リスク比（RR）やリスク差（RD）の統合値と95%信頼区間

# SRの評価 (比較的) 簡単な方法

- 定量的SR (MA) の図表を利用する
  - アウトカムごとに提示される

Bergland score  
type II

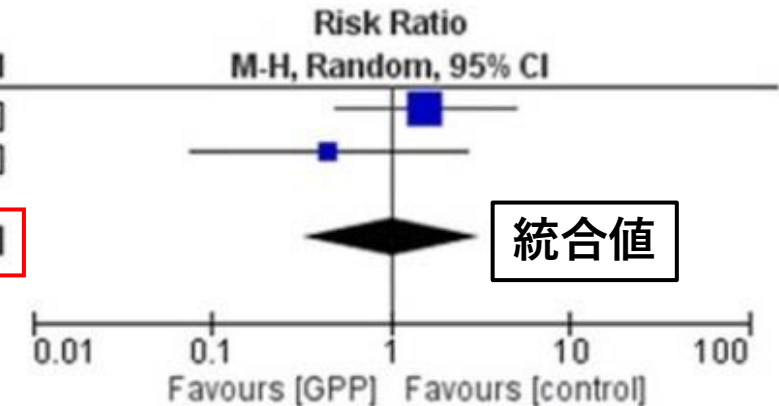
(2b)

Study or Subgroup	GPP		Control		Weight	Risk Ratio M-H, Random, 95% CI
	Events	Total	Events	Total		
Jabbari et al 2017	4	15	5	29	66.1%	1.55 [0.49, 4.92]
Matic and Power 2008	2	20	2	9	33.9%	0.45 [0.07, 2.71]
<b>Total (95% CI)</b>		<b>35</b>		<b>38</b>	<b>100.0%</b>	<b>1.02 [0.32, 3.20]</b>
Total events	6		7			
Heterogeneity: $\tau^2 = 0.17$ ; $\chi^2 = 1.28$ , $df = 1$ ( $P = 0.26$ ); $I^2 = 22\%$						
Test for overall effect: $Z = 0.03$ ( $P = 0.98$ )						

リスク人数

効果指標の統合値  
(この場合はRR)

効果指標の信頼区間



統合値

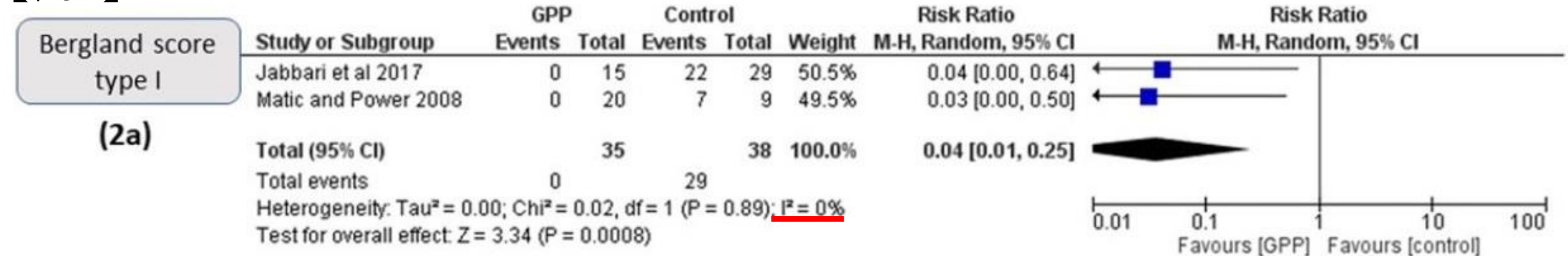
非一貫性の統計学的検定

< 40%	小さい
30~60%	中等度
50~90%	かなり大きい
75~100%	非常に大きい

非一貫性  
不精確性

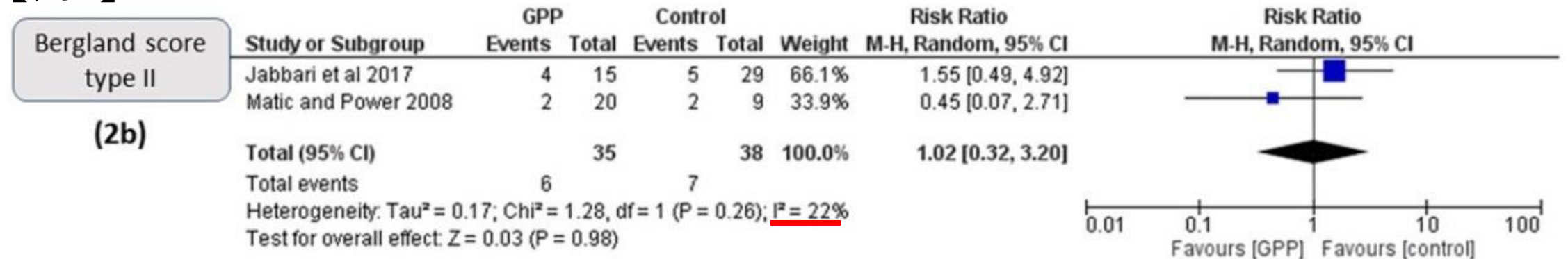
⇒ 各研究のばらつきを見る  
⇒ 統合値と縦中央線の関係を見る

## 【例1】



非一貫性：各研究の信頼区間が概ね一致、 $I^2$ 値 = 0  $\Rightarrow$  なし  
 不精確性：信頼区間の統合値が  $RR = 1$ （中央線）をまたがない  $\Rightarrow$  なし

## 【例2】



非一貫性：各研究の点推定値（■）が  $RR = 1$  の両側、 $I^2$ 値 = 22  $\Rightarrow$  あり（減点）  
 不精確性：信頼区間の統合値が  $RR = 1$  をまたぐ  $\Rightarrow$  あり（減点）



# SRの評価 (比較的) 簡単な方法

- 本文中の記述も参考にする

## *Risk of Bias*

Using the Cochrane tool (ROBINS-I), serious risk of bias was observed in all included studies (Supplementary Table 1). The main reasons were the lack of control on important confounders and the inadequate reporting of the criteria for participant inclusion. Baseline characteristics of the intervention and control groups were only reported in 4 articles (Matic and Power, 2008a; Matic and Power, 2008b; Wang et al., 2016; Jabbari et al., 2017). Assignment of participants to the intervention or control group was done according to the size of the cleft gap in 3 studies, that is, patients with smaller cleft gaps were allocated to the GPP group (Wood et al., 1997; Hsieh et al., 2010; Wang et al., 2016). Nine out of the 13 included articles did not provide any information about the orthodontic treatment performed in any group (Henkel and Gundlach, 1997; Wood et al., 1997; Meazzini et al., 2008; Matic and Power, 2008a; Matic and Power, 2008b; Power and Matic, 2009; Hsieh et al., 2010; Meazzini et al., 2008; Meazzini et al., 2010). Blinding of outcome assessment was reported only in 3 articles (Matic and Power, 2008a; Matic and Power, 2008b; Power and Matic, 2009).

バイアスリスクの程度 (serious) とその理由が記載されている

inconsistency (非一貫性)  
imprecision (不精確性) などの記載

## *Assessment of the Quality of the Evidence*

According to GRADE assessment, the level of evidence was low back to the observational study design of included articles. The GRADE was additionally downgraded due to the following: the increased risk of bias of the included studies (Supplementary Table 1), the inconsistency observed due to high  $I^2$  value, the imprecision of effect estimates caused by the small sample size, and the indirectness of the evidence (Table 3). In fact, reasons that could have upgraded GRADE, such as a large magnitude of intervention effect, the presence of dose-response gradient, and the presence of plausible confounding factors that would work against the direction of the observed treatment effect, were not present.