

ランダム化比較試験 (RCT) 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が…

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

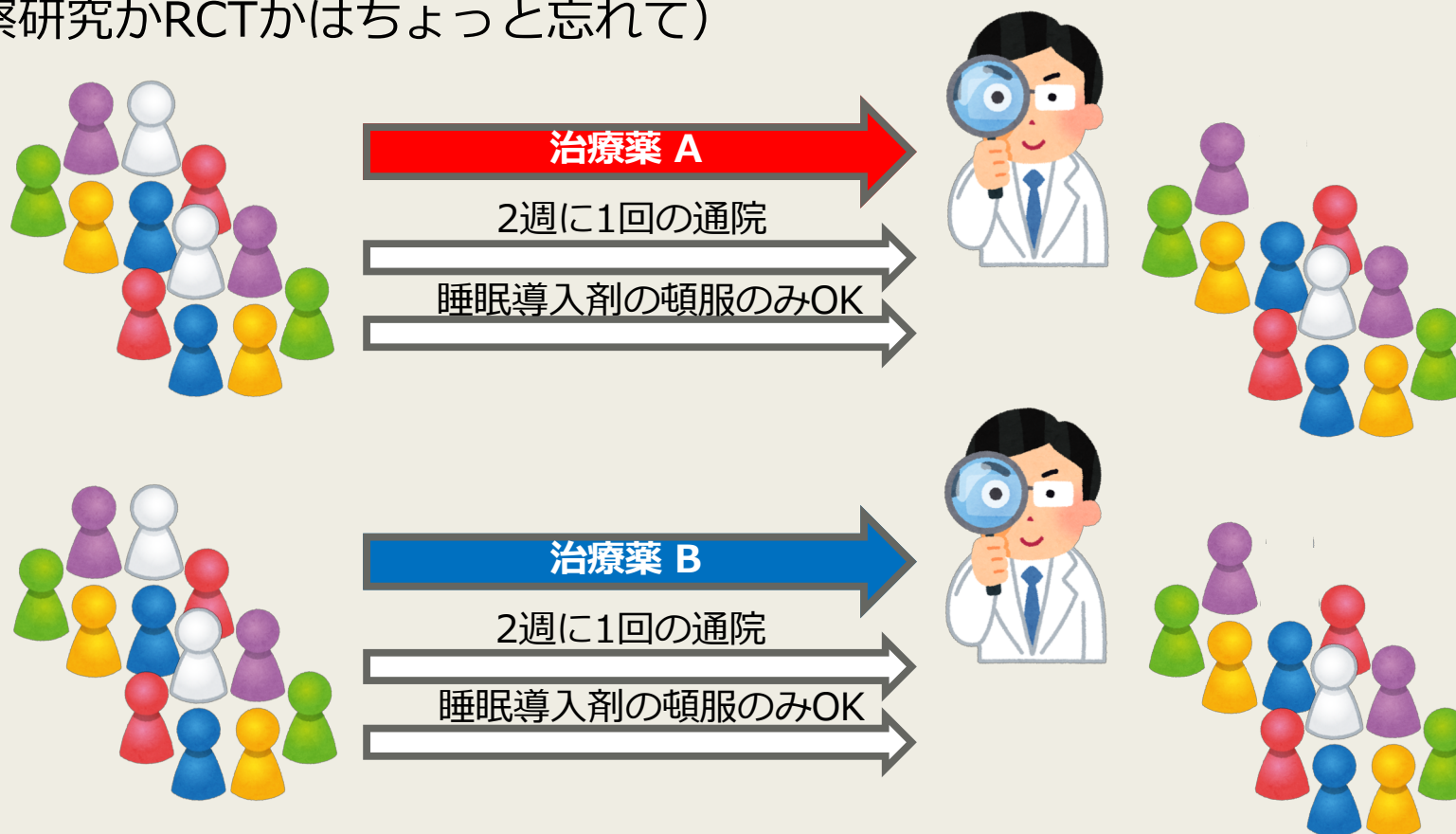
# ベストエビデンスの得られる研究デザインは テーマによって異なります

	観察研究			介入研究
	横断研究	ケース・コントロール研究	コホート研究	RCT
病気や診断の実態	○		○	
病因とアウトカムとの関係	○	◎	◎	○
治療の効果			○	◎
診断法の評価	◎			

※臨床家としては「治療の効果」について調べることが一番多いはず  
※本日は、システマティックレビュー/メタアナリシスについては触れません

# 治療薬Aと治療薬Bの理想的な比較

(観察研究かRCTかはちょっと忘れて)



同等の集団に  
対して

比較したい治療以外は  
同等の治療を行い

同等の方法  
で評価して

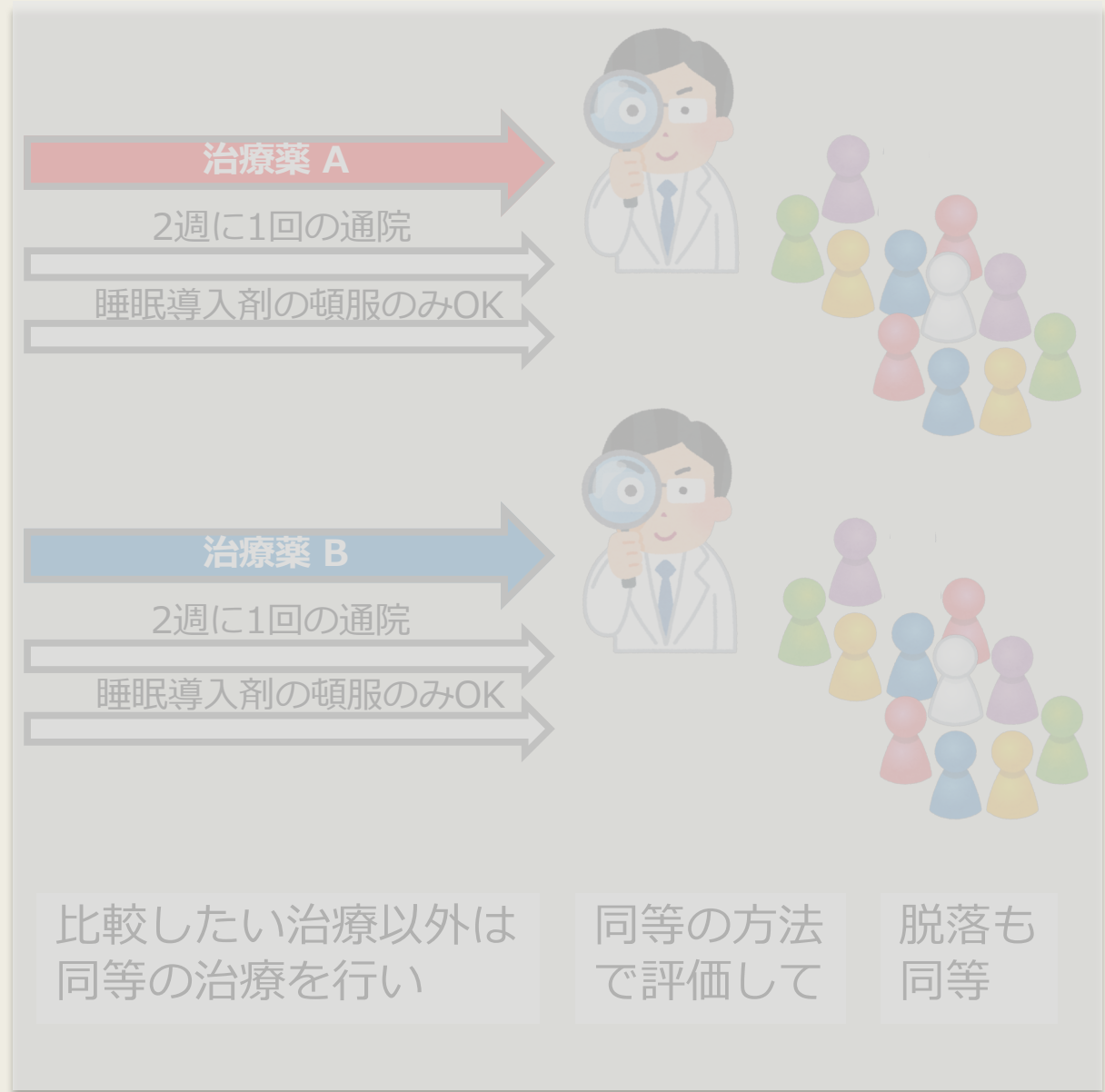
脱落も  
同等



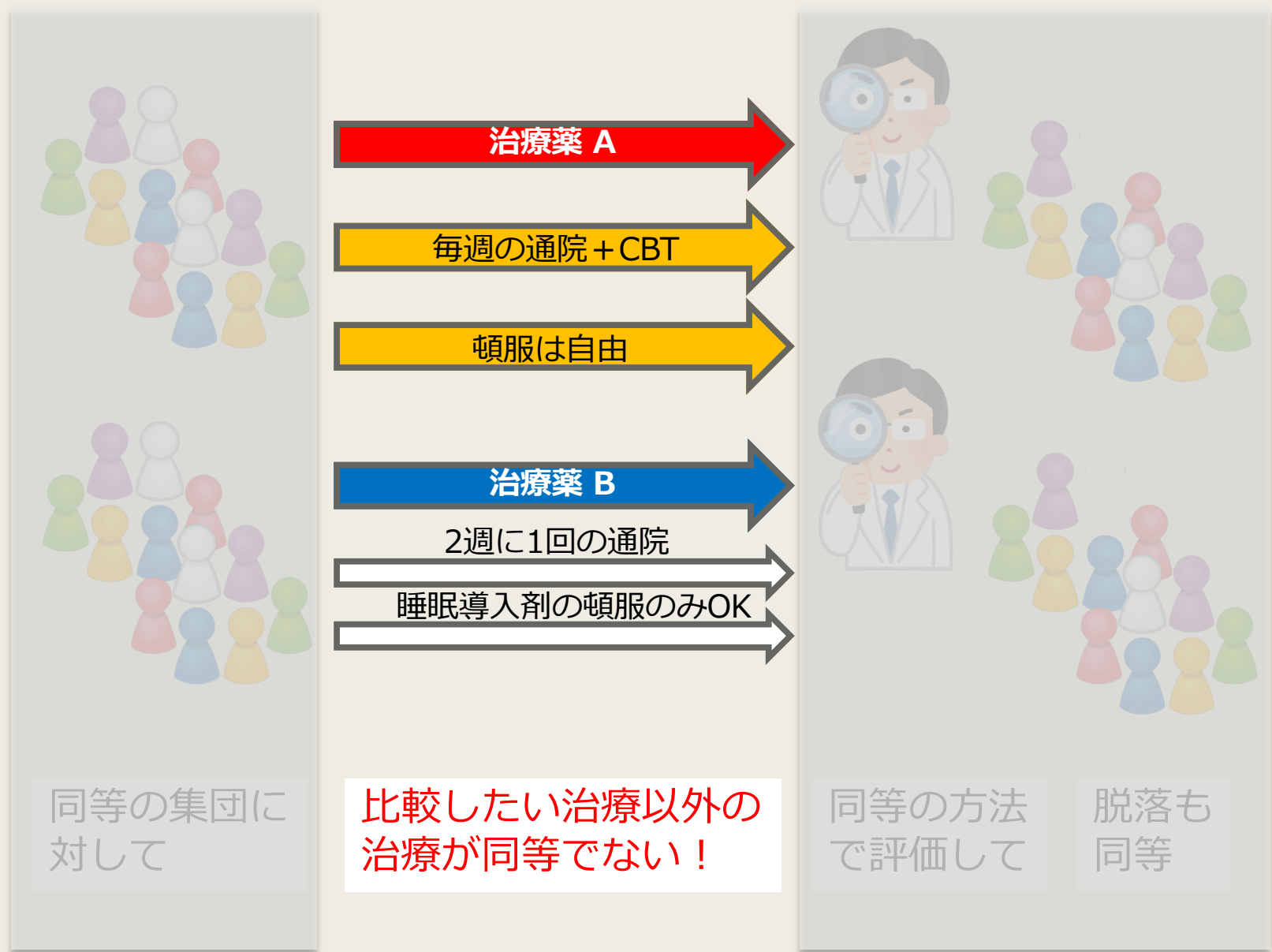
# バイアス：パターン1



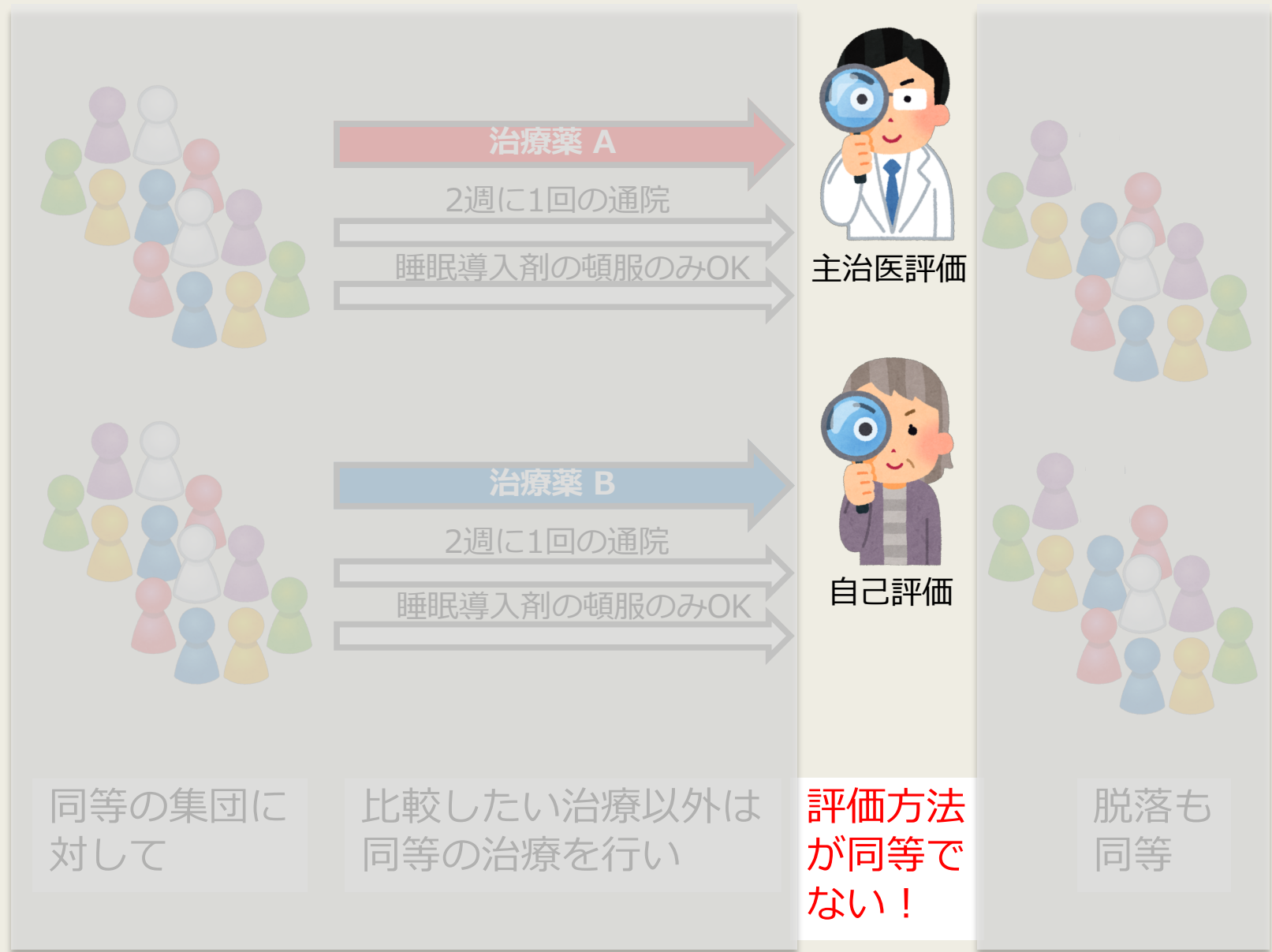
同等の集団とはいえない！



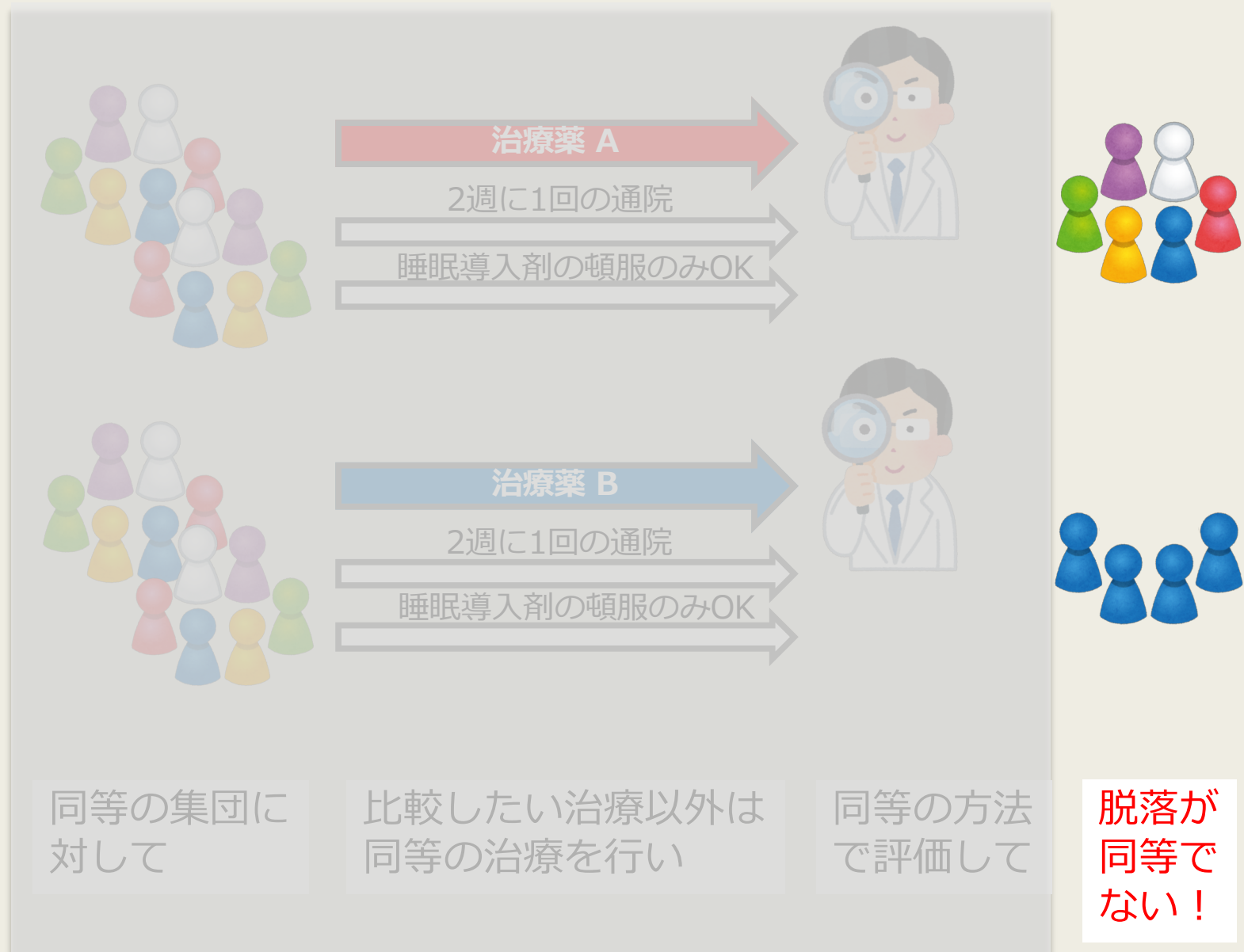
# バイアス：パターン2



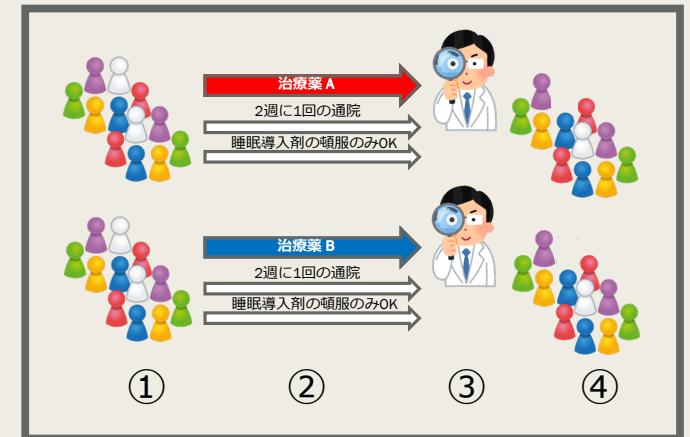
# バイアス：パターン3



# バイアス：パターン4



# 比較の妥当性に影響する 4つのバイアス



## ① 罹病性バイアス (susceptibility bias)

研究開始前の状態において、両群で差があること

## ② 施行バイアス (performance bias)

曝露あるいは介入において、両群で不当な差があること

## ③ 検出バイアス (detection bias)

結果の測定方法に両群で差があること

## ④ 移転バイアス (transfer bias)

脱落などの影響が大きく、両群で差があること

# 観察研究とRCTの違い

## 観察研究



例. 「先生からミルタザピンの方がいいって言われました」

その薬が選ばれたのには**理由**がある  
(両群の背景因子にはおそらく偏りがある)

罹病性バイアスあり

## RCT



例. 「くじでミルタザピンになりました」

その薬が選ばれたのは**偶然**による  
(両群の背景因子は平等に振り分けられる)

罹病性バイアスなし (のはず)



さらに...



# 観察研究とRCTの違い



さらに...



## 観察研究

治療が進んでいく自然経過を  
観察しているだけなので

- ・ 施行バイアス
- ・ 検出バイアス
- ・ 移転バイアス

を制御できない



比較の妥当性は  
担保されていない

## RCT

実験的な条件下で治療経過を  
追跡している

- ・ 施行バイアス
- ・ 検出バイアス
- ・ 移転バイアス

を制御できている (はず)



比較の妥当性は  
担保されている (はず)

# 観察研究は比較の妥当性が担保されないので…

- 「観察研究では治療効果を過大に評価してしまう」

Grimes. Obstet Gynecol. 2012

Ioannidis. Epidemiology. 2008

Ioannidis. JAMA. 2001

- 「コホート研究ではリスク比が2~3以上, ケース・コントロール研究ではオッズ比が3~4以上はないと, その効果はバイアスに基づくものである可能性が高い」

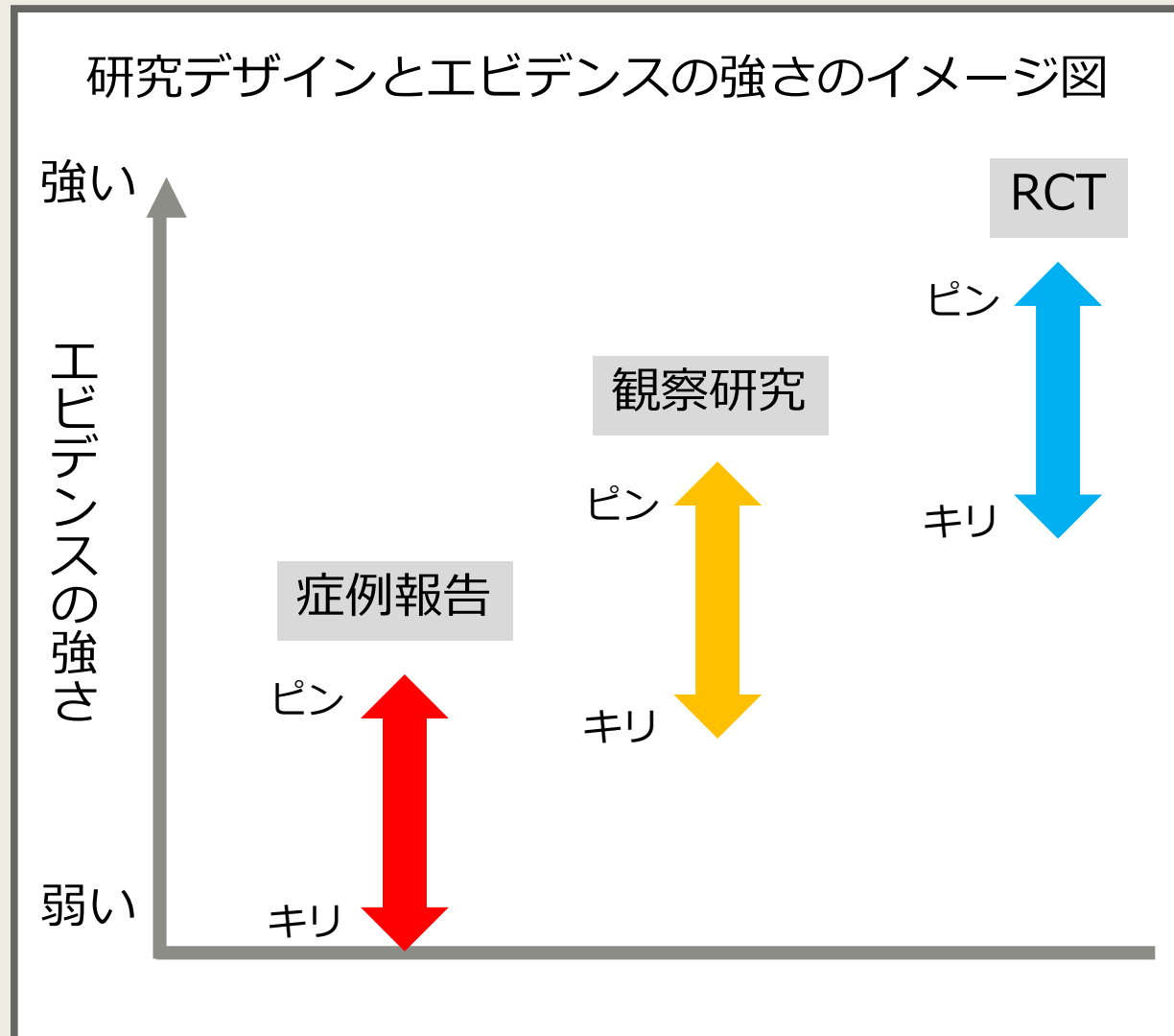
Grimes. Obstet Gynecol. 2012

観察研究の結果のみから治療の有効性を述べている場合,  
のちにその結果がRCTによって覆される可能性は高い

治療法に関する論文を読むなら, やはりRCTの中から選ぶべき



# じゃあ、RCTならすべてOK？



すべての研究は  
ピンからキリまでである

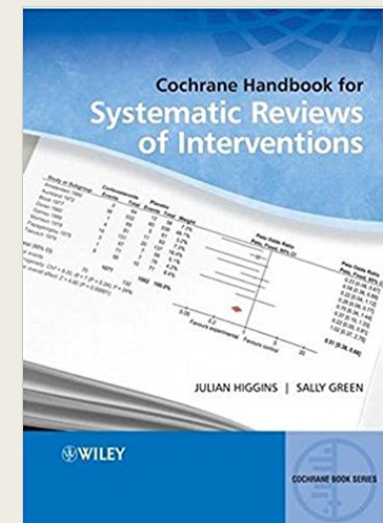
論理的には、RCTが  
他のデザインに優って  
いるのは事実だが、  
個々のRCTもピンキリ



批判的にチェック  
しないと痛い目にあう

# では、RCT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、メタアナリシスやガイドライン作成の際には必須となる知識です



# 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

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

治療に関する論文の批判的吟味

内的妥当性

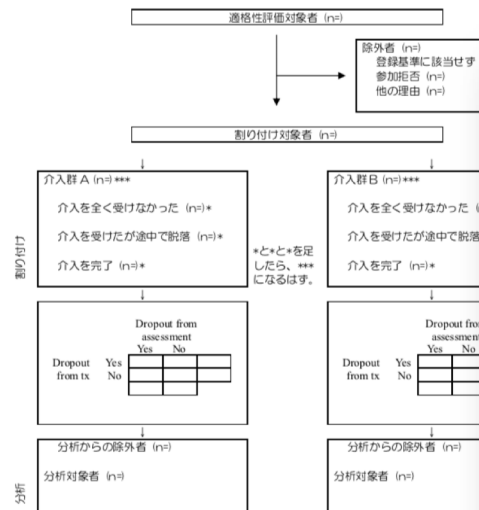
(以下の項目について、○、△、×で評価し、その理由を書いてください)

1. 治療への割り付けは無作為に行われていたか、割り付けは隠蔽化されていたか	評価 : ○    △    ×
コメント :	
2. 臨床試験にエントリーされた全ての患者が試験の終了時点でどうなったかについて説明	評価 : ○    △    ×
コメント :	
3. 患者、治療者、評価者、データ分析者は、割り付けられた治療についてブラインドであ	評価 : ○    △    ×
コメント :	
4. 治療群と対照群の間で、臨床試験の開始時に重要な臨床特性において差はなかったか	評価 : ○    △    ×
コメント :	
5. 治療群と対照群は臨床試験の目的である介入以外では、同等に扱われたか	評価 : ○    △    ×
コメント :	

自由記載欄 :

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

論文のサマリ



結果

- ① Good event または Bad 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%信頼区間はどうか。[<http://ebmh.med.kyoto-u.ac.jp/toolbox/>]

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	イベント率
介入群 A			
介入群 B			

95%

OR=		下限	上限
RR=			
NNT=			

95%

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  
2/4

③論文の結果を自身の症例に当てはめて考える

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?

※ 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

治療に関する論文の批判的吟味

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)

1. 治療への割り付けは無作為に行われていたか、割り付けは隠蔽化されていたか	
評価	: ○      △      ×
コメント:	罹病性バイアス
2. 臨床試験にエントリーされた全ての患者が試験の終了時点でどうなったかについて説明があるか	
評価	: ○      △      ×
コメント:	移転バイアス
3. 患者、治療者、評価者、データ分析者は、割り付けられた治療についてブラインドであったか	
評価	: ○      △      ×
コメント:	検出バイアス, 施行バイアス
4. 治療群と対照群の間で、臨床試験の開始時に重要な臨床特性において差はなかったか	
評価	: ○      △      ×
コメント:	罹病性バイアス
5. 治療群と対照群は臨床試験の目的である介入以外では、同等に扱われたか	
評価	: ○      △      ×
コメント:	施行バイアス

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

1. 治療への割り付けは無作為に行われていたか  
 割り付けは隠蔽化されていたか  
 (罹病性バイアス)

- ①割り付けのためのランダムな文字列が、コンピュータプログラムや乱数表等できちんと作られていたか
- ②さらに、割り付けをする人に恣意的な操作ができないよう、それがきちんと隠蔽化されていたか



← ②壺振りがいかさまをする余地がなかったか？

← ①いかさまサイコロを使っていなかったか？

この両者を分けて考える

## 2. 臨床試験にエントリーされた全ての患者が試験の終了時点でどうなったかについて説明があるか (移転バイアス)

- 論文のFigure1あたりに脱落者の脱落理由が書いてあるか
- 「治療からの脱落」と「評価からの脱落」を分けて考える  
治療からの脱落はやむを得ないが、評価は継続するべき
- 割り付けられた全員を解析対象者とするのがRCTの大原則  
**ITT (intention to treat) 分析** ... だが、現実はなかなかそうはなっていない
- 適切な方法で欠測値が補完されているか
- 行方不明 (lost to follow-up) が一番の問題  
良くなって脱落したのか、悪くなって脱落したのかで話は大きく変わってきてしまう

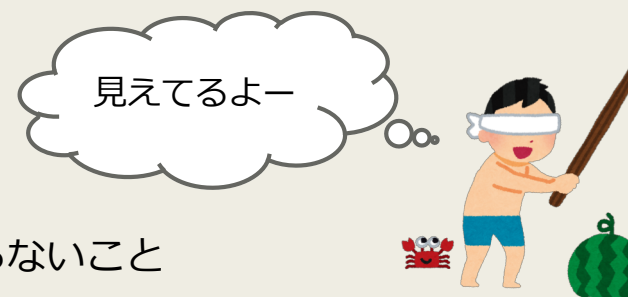


### 3. 患者，治療者，評価者，データ分析者は，割り付けられた治療についてブラインドであったか (検出バイアス，施行バイアス)

- 「患者」「治療者」「評価者」「データ分析者」に分けて評価
- アウトカムを評価しているのは誰か，患者の自己評価なのか，治療者が評価しているのか，あるいは独立に評価者を設定しているのか
- 精神療法などは，患者と治療者をブラインドにできないため，独立な評価者を立ててブラインドにするべき
- ブラインドが破られている場合，効果の過大評価につながる

注. ブラインド ≠ 割り付けの隠蔽化

ブラインドとは，割り付けられた治療が何かわからないこと  
(割り付けが済んだ後の話)

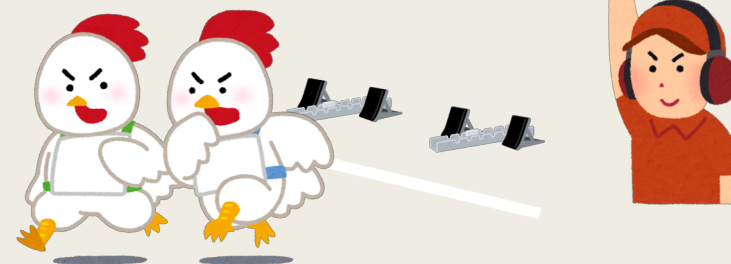




#### 4. 治療群と対照群の間で，臨床試験の開始時に重要な臨床特性において差はなかったか (罹病性バイアス)

- ランダム割り付けの方法に問題がなかったとしても，スタート時点の臨床特性に，両群で偶然の偏りが生じる可能性はある
- 論文のTable1を参照
- 単に統計学的な差があったかということより，結果に影響を及ぼす重要な特性に差がなかったかを“臨床的に”考える

「性別」「年齢」「婚姻状況」「抑うつ重症度」…  
何が結果に影響するかは，状況によって異なる

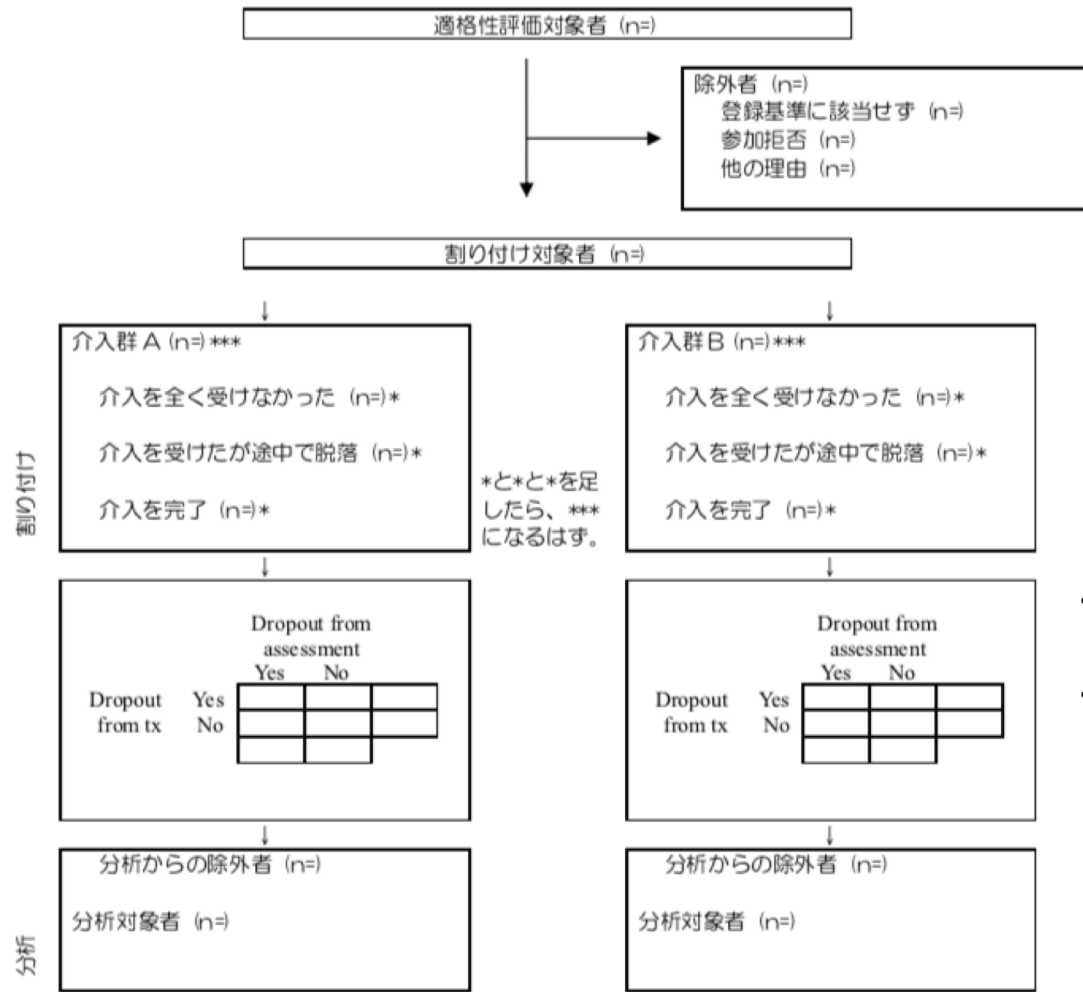


## 5. 治療群と対照群は臨床試験の目的である介入以外では、同等に扱われたか (施行バイアス)

- 治療群と対照群で、併用療法（共介入）や通院頻度などに違いがあるために、どちらかが有利あるいは不利になっていなかったか
- 論文のMethodsに、許容される併用薬等についての記載があり、Resultsに実際にどのような併用薬が使われたかの記載がある（はずだが、記載がないことも多い…）



論文のサマリ



論文のFigure1を参考に埋めていく

- Dropout from treatment (治療からの脱落)
  - Dropout from assessment (評価からの脱落)
- を分けて、2×2表で考える

治療から脱落しても評価は継続するべき



割り付けられた全員を分析対象者とするのがITT (intention to treat) の大原則  
 何らかの理由で分析から除外されたケースがあれば、その数をチェック  
 →分母が減れば、その分、効果の過大評価につながる可能性がある！

結果

- ① Good event または Bad 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%信頼区間はどうか。[http://ebmh.med.kyoto-u.ac.jp/toolbox.html 参照]

単純な ITT 分析

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

95%

	下限	上限
OR=		
RR=		
NNT=		

Worst case scenario

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

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。有意差がある場合は worst case

2/4

教室ホームページから「NNT Calculator」のエクセルシートをダウンロード  
A群, B群, それぞれ何人中何人にイベントが発生したかを入力すると, 自動的にNNTが計算される

単純なITT分析で有意差があれば (=有効性が認められれば) ワーストケースシナリオでも計算する  
(コントロール群の脱落者が, 実は全員良くなっていたら...という極端な仮定)  
それでも有意差があれば, その治療が有効であることの信憑性は, より高まる

注. "good event"か"bad event"かによって  
考え方は反対になるが, 本質は同じ



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、この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
- もしこの研究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. 論文の包含基準が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RCT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