

メタアナリシスの方法と実践

手良 向 聡

京都大学医学部附属病院 探索医療センター検証部

京都大学大学院医学研究科 臨床試験管理学分野

ICR臨床研究入門中級編、2009.10.28

1

参考書(医学分野)

(入門)

- 手良向聡. メタアナリシス. In: 臨床試験ハンドブック, 朝倉書店, 2006:633-655.
- 丹後俊郎. メタ・アナリシス入門. 朝倉書店, 2002.
- Borenstein M, et al. Introduction to Meta-Analysis. Wiley, 2009.
- Sutton AJ, et al. Methods for Meta-Analysis in Medical Research, Wiley, 2000.

(応用:臨床試験)

- Whitehead A. Meta-Analysis of Controlled Clinical Trials, Wiley, 2002.
- Egger M, et al., eds. Systematic Reviews in Health Care: Meta-Analysis in Context, BMJ Publishing, 2001.

(応用:観察研究)

- Greenland S, et al. Meta-Analysis. In: Modern Epidemiology, 3rd ed. Lippincott Williams & Wilkins, 2008:652-682.

メタアナリシスとは？

- Systematic review (系統的レビュー)
 - 質の高い証拠を系統的に収集して行う総合的評価
- Meta-analysis (Glass GV, 1976)
 - The statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings
 - 知見の統合を目的とした複数の研究結果の統計解析
- 同義語
 - overview、quantitative review、synthesis、pooling、combining、integration

メタアナリシスの歴史

- 1976～: 当初主に米国において教育学、心理学などの社会科学分野で応用
- 1977: 医学分野において最初のメタアナリシス(急性心筋梗塞、Chalmers TC)
- 1985: 心筋梗塞に対する β ブロッカーのメタアナリシス(Yusuf S, Peto R)
- メタアナリシスに関する報告数(PubMed)
 - 1990年は約300報、2002年は1000報を超え、2005年には2000報を超える
- 方法論に関する報告数(Stat Med)
 - 1995年～2002年に集中、その後やや停滞

メタアナリシスの目的

- 特定の研究領域からの証拠を系統的に概観すること
- 各研究から得られる結果の定量的な要約を提供すること
- 適切であれば、研究にわたって結果を併合すること
- 研究間のばらつきの量を評価すること
- 研究間のばらつきに関連する特性を同定すること

留意点

- **メタアナリシスの目的**
 - 個々の臨床研究と同様に、ある臨床的な仮説に対して現時点での解答を与えることであり、仮説の妥当性の吟味および種々のバイアスに対する考慮がメタアナリシスの意義および質を決定する
- **メタアナリシスの方法論**
 - 研究のタイプ(臨床試験、観察研究など)に依存しないが、観察研究は臨床試験に比べて多くのバイアスを含んでいるため、観察研究のメタアナリシスには結果の解釈により多くの注意が必要である
- **個々の研究の質が、メタアナリシスの質を決定する**

実施手順

- 通常は後向き（retrospective）に実施、前向き（prospective）なメタアナリシスも可能である
- 目的、試験の適格規準、エンドポイント、統計学的考察を含むプロトコル（研究実施計画書）を作成する
 - ただし、適格規準に合致した試験が存在しない、期待したエンドポイントで結果が報告されていない、データ不備のため計画した統計解析手法が利用できないなどの理由で必ずしもプロトコル通りに実施できない場合がある

試験の存在確認と選択

- MEDLINE、Cochrane Libraryなどの電子化された文献データベースを検索する
- 関連する文献を抽出し、それらの参考文献、会議録など他の情報源からも網羅的に適格規準に合致した試験を収集する
- 収集された試験の質を評価するために、CONSORT*声明などを参考にして、チェックリストを作成する
- 評価の信頼性を上げるために複数の評価者が独立に評価を行う

*: Consolidated Standards of Reporting Trials

データの収集と情報の抽出

- 文献に基づくメタアナリシス
 - あらかじめ定めたエンドポイントおよび患者特性などに関する情報を文献から正確に抽出しなければならない
- 個人データ (IPD: individual patient data) に基づくメタアナリシス
 - 文献に基づく場合に比べて多くの利点を有することは明らかであるが、データ収集には時間と費用が掛かり全適格試験のデータ収集が困難な場合もある

解析と報告

- 臨床試験のエンドポイントに対応する効果の尺度を特定した上で、結果の異質性の評価、結果の統合および異質性を説明する要因の探索などを行う
- メタアナリシスの報告については、QUOROM*声明が参考になる

*: the Quality of Reporting of Meta-analyses

事例1: 進行非小細胞肺癌に対する化学療法

- 2,948名(8試験)
- 文献に基づく
- プラチナ製剤(シスプラチンベース併用療法とカルボプラチンベース併用療法)の比較
- 死亡ハザード比(シスプラチンを基準)
 - 全試験 HR: 1.05 (95%CI, 0.91-1.22)
 - 新規薬との併用(5試験) HR: 1.11 (95%CI, 1.01-1.22)
- 安全性
 - シスプラチンベースの方が悪心嘔吐多い
 - カルボプラチンベースの方が血小板減少多い

Hotta K, et al. J Clin Oncol 2004;22:3852-3859

事例2: 糖尿病治療薬と心血管イベント

- 27,843名(42試験)
- 文献に基づく
- ロシグリタゾン(アバンディア)の評価
- オッズ比(対照群を基準)
 - 心筋梗塞発生 OR: 1.43 (95%CI, 1.03-1.98)
 - 心血管イベント死亡 OR: 1.64 (95%CI, 0.98-2.74)
- 問題点
 - まれなイベント(発生割合:心筋梗塞発生0.6%、心血管イベント死亡0.2%)の場合、解析手法によって結果が大きく異なる

Nissen SE, et al. N Engl J Med 2007;356:2457-2471

事例3: コレステロール低下薬(スタチン)の効果

- 90,056名(14試験)
- 個人データに基づく
- 前向き
 - メタアナリシスのプロトコル公表(1995)
 - 各試験結果の公表(1994-2004)
- スタチン(5種類)の評価
- リスク比(追跡1年時のLDL-C1mmol/L(39mg/dL)低下あたり)
 - 虚血性心疾患死亡 RR: 0.81 (95%CI, 0.76-0.85)
 - 総死亡 RR: 0.88 (95%CI, 0.84-0.91)
 - 癌の発生 RR: 1.00 (95%CI, 0.95-1.06)

Cholesterol Treatment Trialists' (CTT) Collaborators. Lancet 2005;366:1267-1278

効果の尺度

- 相対尺度 (relative measure)
 - オッズ比 (odds ratio) ……二値分類型
 - リスク比 (risk ratio) ……二値分類型
 - ハザード比 (hazard ratio) ……時間-イベント型
- 絶対尺度 (absolute measure)
 - リスク差 (risk difference) ……二値分類型
 - 平均値の差 (mean difference) ……連続型
 - 標準化した平均値の差 (standardized mean difference)

オッズ比

群	イベント発生		計
	有	無	
治療群	a	b	n ₁
対照群	c	d	n ₀
計	m ₁	m ₀	n

$$OR = \frac{a/b}{c/d} = \frac{ad}{bc} \quad (1)$$

リスク比

群	イベント発生		計
	有	無	
治療群	a	b	n_1
対照群	c	d	n_0
計	m_1	m_0	n

$$RR = \frac{a/n_1}{c/n_0} \quad (2)$$

ハザード比

- 集団でのハザード率 (hazard rate)
 - 発生率 (incidence rate)
= イベント発生数 / 観察人・時間
- ハザード比
 - 治療群のハザード率 / 対照群のハザード率

リスク差

群	イベント発生		計
	有	無	
治療群	a	b	n_1
対照群	c	d	n_0
計	m_1	m_0	n

$$RD = \frac{a}{n_1} - \frac{c}{n_0} \quad (3)$$

平均値の差

- 平均値の差

$$MD = \bar{X}_1 - \bar{X}_0 \quad (4)$$

- 標準化した平均値の差

$$SMD = \frac{\bar{X}_1 - \bar{X}_0}{s} \quad (5)$$

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_0 - 1)s_0^2}{n_1 + n_0 - 2} \quad (6)$$

解析手法

- 文献に基づくメタアナリシス -

表1. 臨床試験 i の結果の要約

群	イベント発生		計
	有	無	
治療群	a_i	b_i	n_{1i}
対照群	c_i	d_i	n_{0i}
計	m_{1i}	m_{0i}	n_i

結果の異質性

- メタアナリシスにおける結果のばらつきには次の3つの要素がある
 1. 試験内のばらつき
 - 試験ごとの治療効果の推定に伴うばらつきで、通常、純粹にランダムな誤差と考える
 2. 試験間の治療効果のばらつき
 - 一般に結果の異質性と呼ばれ、効果の大きさ(effect size)が試験間で一様でないことを意味する
 3. 対照群リスクのばらつき
 - 対照群における結果(例えば、死亡割合、ハザード率など)が試験間で異なることを意味する

異質性の検定・評価

- 試験*i*における効果の大きさの推定値を T_i ($i=1, \dots, k$)、試験*i*に与えられる重みを w_i とし、効果の大きさの重み付き推定値を

$$\bar{T} = \frac{\sum_{i=1}^k w_i T_i}{\sum_{i=1}^k w_i} \quad (7)$$

とする。このとき、結果の異質性に関する検定は、

$$Q = \sum_{i=1}^k w_i (T_i - \bar{T})^2 \quad (8)$$

という統計量に基づいて行える

- 各試験の効果の大きさが全て等しいという帰無仮説の下で、 Q 統計量は自由度 $k-1$ の χ^2 分布に従う

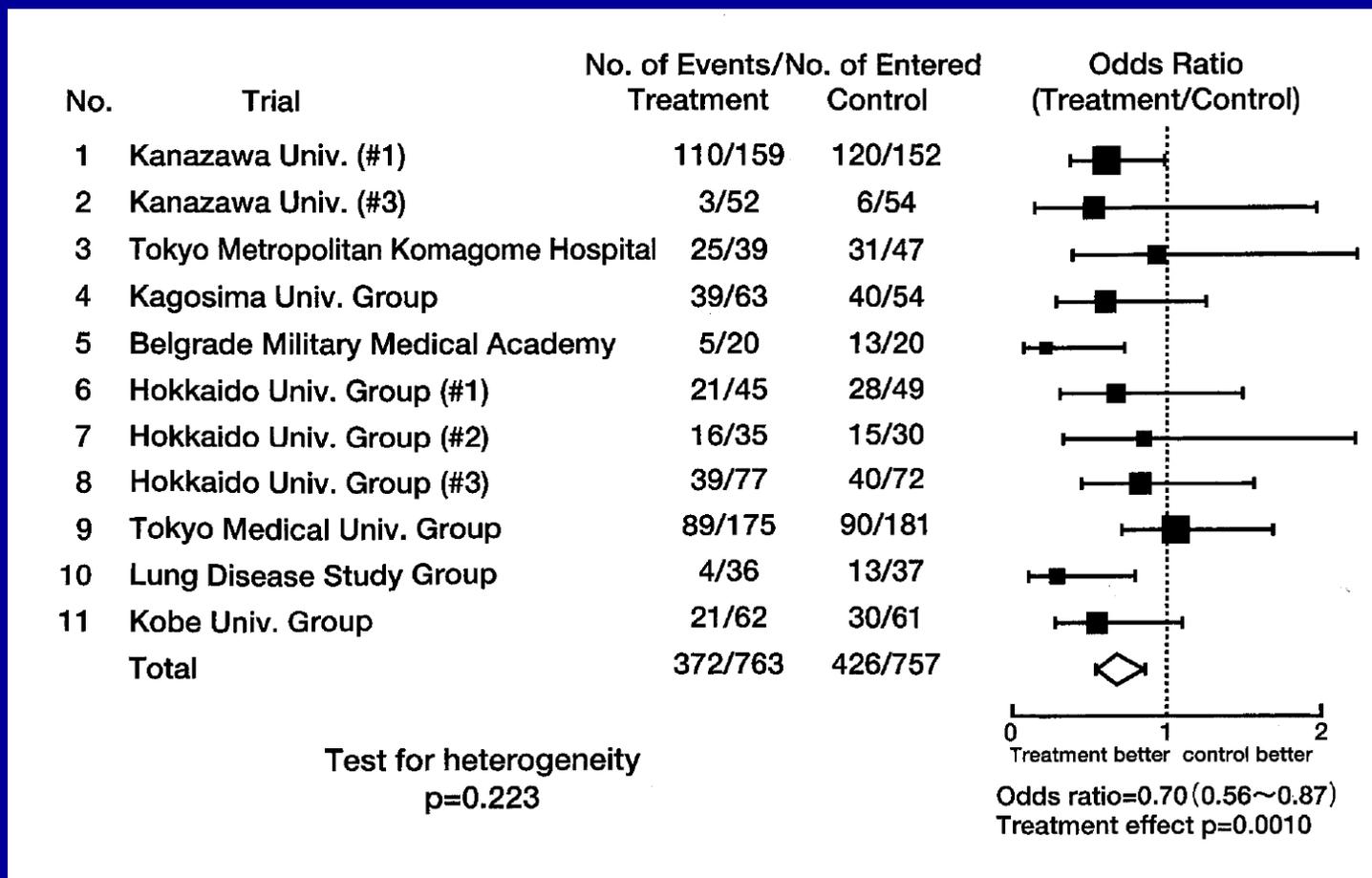
- 試験数が少ない場合、この検定の統計的検出力は非常に低いことが知られており、有意水準として0.10を用いることが推奨されている
- 一方、各試験の対象数が非常に多い場合には、帰無仮説を棄却しやすい傾向がある
- 異質性の程度を定量的に評価する指標として以下の I^2 が提案されている

$$I^2 = \frac{Q - (k - 1)}{Q} \times 100\% \approx \frac{\text{試験間分散}}{\text{試験間分散} + \text{試験内分散}} \times 100\% \quad (10)$$

- 0%～100%の値で、0%に近い場合は結果の一致度が高く、100%に近い場合は結果の一致度が低い
- 式(8)のQ統計量と試験数 k のみから算出できるため、公表論文から結果の異質性を評価する際にも有用である

グラフ表示

図1. 森林プロット (forest plot) / 信頼区間プロットの例



Sakamoto J, Teramukai S, Watanabe Y, et al. J Immunother 2001;24:250-256

25

結果の統合

- まず各試験で得られた結果を**同一の効果の尺度**を用いて表す
- その尺度で表された各試験の効果の大きさをある方法で統合して**統合推定値 (pooled estimate)**を得る
- 代表的な統計モデル
 - **母数(固定)効果モデル (fixed-effects model)**
 - 試験間の治療効果のバラツキを偶然誤差と扱う
 - **変量効果モデル (random-effects model)**
 - 試験間の治療効果のバラツキ(異質性)を考慮して、各試験をある集団からのサンプル、すなわち変量として扱う
 - 確率分布のパラメータ(超パラメータ)の不確実性を考慮した**ベイズ流モデル (Bayesian model)**に拡張できる

母数効果モデル

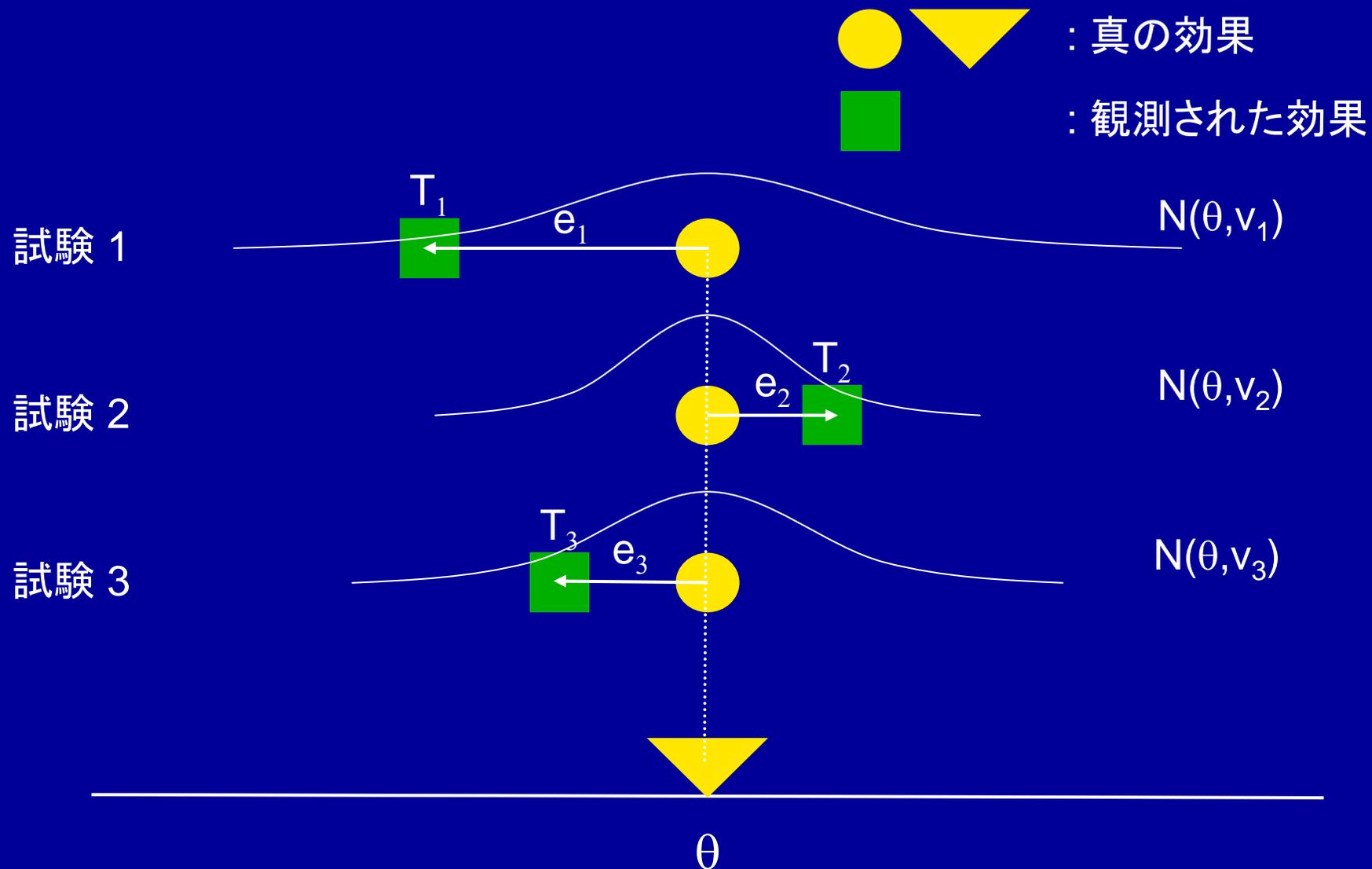
- 各試験の真の効果の大きさは同一、すなわち、真の各試験共通の効果の大きさを θ 、各試験の真の効果の大きさを $\theta_i (i = 1, \dots, k)$ としたとき、 $\theta_1 = \dots = \theta_k = \theta$ と仮定する
- 試験*i*における効果の大きさの推定値を T_i とすると、

$$T_i = \theta + e_i \quad (11)$$

ここで、 e_i は標本誤差である

- 統合推定値を得る考え方は、試験を層とした層別解析 (stratified analysis)と同じであり、**逆分散法** (inverse variance-weighted method)、**Mantel-Haenszel法**、**Peto法**などが代表的な方法である

母数効果モデル



逆分散法

- 各試験の効果の大きさを、その分散の逆数を重みとして統合するという単純な方法であり、この原理はあらゆる効果の尺度に対して適用可能である
- 治療効果の統合推定値は、

$$\bar{T} = \frac{\sum_{i=1}^k w_i T_i}{\sum_{i=1}^k w_i} \quad (12)$$

- この統合推定値の分散を最小にする重みは、 T_i の分散 v_i の逆数に比例することが知られており、重み w_i として、

$$w_i = \frac{1}{v_i} \quad (13)$$

- この統合推定値の分散は重みの和の逆数、

$$\text{Var}(\bar{T}) = \frac{1}{\sum_{i=1}^k w_i} \quad (14)$$

- この統合推定値が正規分布に従うと仮定すると、 θ の95%信頼区間は、

$$\bar{T} - 1.96 \sqrt{\frac{1}{\sum_{i=1}^k w_i}} \leq \theta \leq \bar{T} + 1.96 \sqrt{\frac{1}{\sum_{i=1}^k w_i}} \quad (15)$$

- 治療効果がないという帰無仮説の下で、

$$\bar{T}^2 / \text{Var}(\bar{T})$$

は自由度1の χ^2 分布に従う

変量効果モデル

- 各試験の効果の大きさに潜在的な異質性があることを仮定する
- 試験*i*における効果の大きさの推定値を T_i 、真の効果の大きさを θ_i 、標本誤差を e_i とすると、

$$T_i = \theta_i + e_i \quad (28)$$

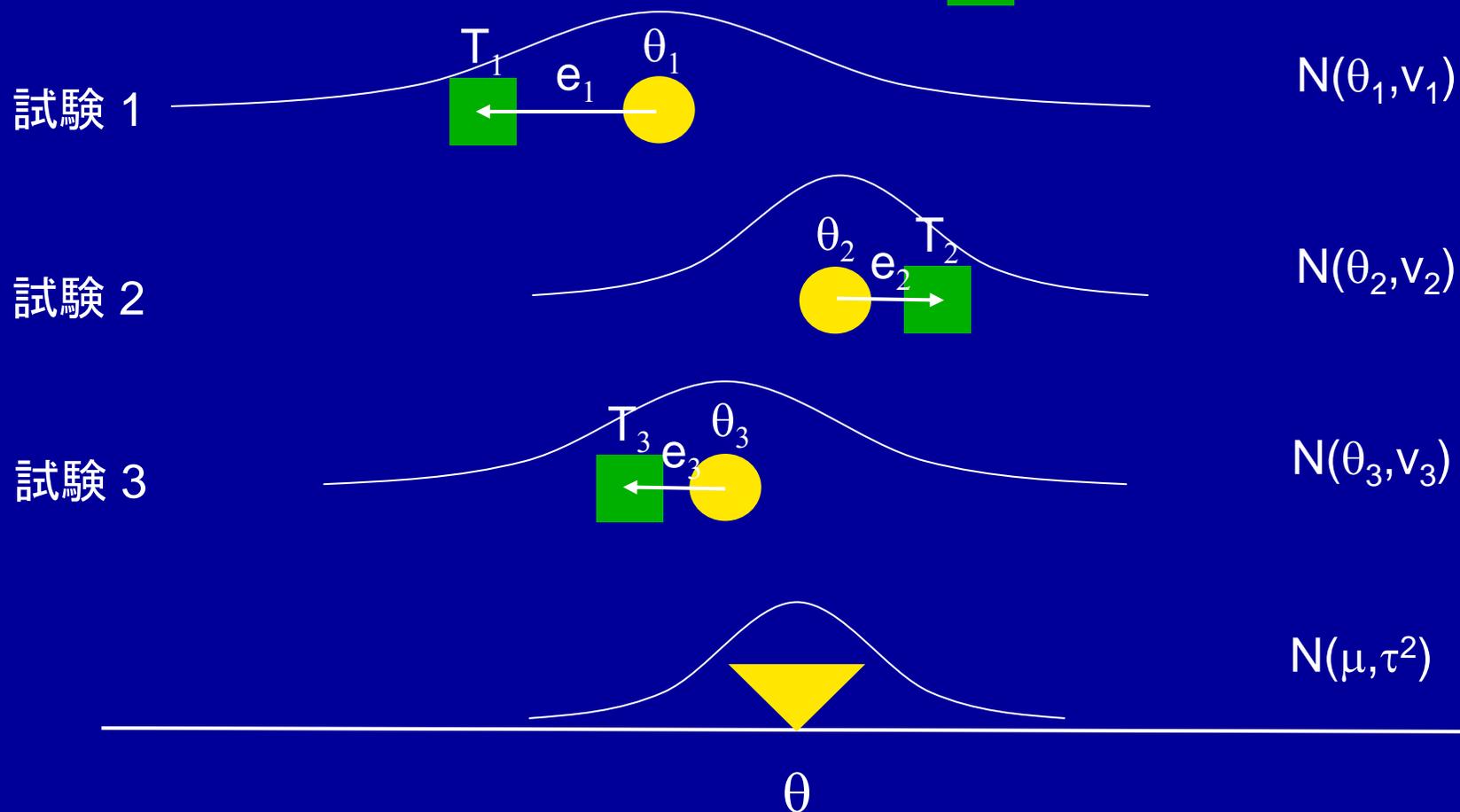
$$\text{Var}(T_i) = \tau^2 + v_i \quad (29)$$

ここで、 τ^2 は効果の大きさの試験間分散、 v_i は試験内分散を表す

- 試験間分散 τ^2 の推定には、1段階モーメント法(通称、**DerSimonian-Laird法**)がその簡便性から最も利用されている

変量効果モデル

● ▼ : 真の効果
■ : 観測された効果



DerSimonian-Laird法

- 試験*i*の重み w_i の平均と分散を

$$\bar{w} = \frac{1}{k} \sum_{i=1}^k w_i \quad (30)$$

$$s_w^2 = \frac{1}{k-1} \left(\sum_{i=1}^k w_i^2 - k\bar{w}^2 \right) \quad (31)$$

- Qを式(8)で定義した統計量、

$$U = (k-1) \left(\bar{w} - \frac{s_w^2}{k\bar{w}} \right) \quad (32)$$

として、 τ^2 を以下の式で推定する

$$\hat{\tau}^2 = \max \left\{ 0, \frac{Q - (k-1)}{U} \right\} \quad (33)$$

- 試験*i*の重み w_i^* は試験内と試験間の分散の和の逆数、

$$w_i^* = \frac{1}{\hat{\tau}^2 + v_i} \quad (34)$$

- この重みを用いて、統合推定値、その分散、その95%信頼区間は、

$$\bar{T} = \frac{\sum_{i=1}^k w_i^* T_i}{\sum_{i=1}^k w_i^*} \quad (35)$$

$$\text{Var}(\bar{T}) = \frac{1}{\sum_{i=1}^k w_i^*} \quad (36)$$

$$\bar{T} - 1.96 \sqrt{\frac{1}{\sum_{i=1}^k w_i^*}} \leq \theta \leq \bar{T} + 1.96 \sqrt{\frac{1}{\sum_{i=1}^k w_i^*}} \quad (37)$$

どちらのモデルを用いるか？

- すべての試験がほぼ同一とみなせる対象で、**その集団について(他の集団に外挿できない)共通の効果の大きさを計算したいとき**
 - 母数効果モデル
- それ以外の多くの場合
 - 変量効果モデル
- まず母数効果モデルで解析して、異質性の検定結果が有意なときに、変量効果モデルに移行するというデータに基づく手順は推奨できない

解析例

臨床研究・生物統計研究会誌 2003;23:83-89

表2. 結果の要約: 非小細胞肺癌データ

試験番号	死亡数／対象数 (死亡割合)		オッズ比	95%信頼区間
	治療群	対照群		
1	110/159 (69%)	120/152 (79%)	0.60	0.36～1.00
2	3/52 (6%)	6/54 (11%)	0.49	0.12～2.07
3	25/39 (64%)	31/47 (66%)	0.92	0.38～2.24
4	39/63 (62%)	40/54 (74%)	0.57	0.26～1.26
5	5/20 (25%)	13/20 (65%)	0.18	0.05～0.70
6	21/45 (47%)	28/49 (57%)	0.66	0.29～1.48
7	16/35 (46%)	15/30 (50%)	0.84	0.32～2.24
8	39/77 (51%)	40/72 (56%)	0.82	0.43～1.56
9	89/175 (51%)	90/181 (50%)	1.05	0.69～1.59
10	4/36 (11%)	13/37 (35%)	0.23	0.07～0.80
11	21/62 (34%)	30/61 (49%)	0.53	0.26～1.10
要約	372/763	426/757	0.70	0.56～0.87

Sakamoto J, Teramukai S, et al. J Immunother 2001;24:250-256.

37

表3. 逆分散法とDerSimonian-Laird法の相対的重み

試験番号	症例数	リスク比	相対的重み (%)	
			逆分散法	DerSimonian-Laird 法
1	311	0.88	41.9	29.0
2	106	0.52	0.4	0.6
3	86	0.97	7.5	9.7
4	117	0.84	11.7	13.7
5	40	0.39	1.1	1.7
6	94	0.82	4.7	6.5
7	65	0.91	2.8	4.1
8	149	0.91	8.0	10.2
9	356	1.02	17.2	17.8
10	73	0.32	0.7	1.1
11	123	0.69	3.9	5.6
統合リスク比			0.88	0.87
(95%信頼区間)			(0.81-0.96)	(0.78-0.97)

表4. モデル・方法別のメタアナリシス結果の要約

モデル	方法	オッズ比	95%信頼区間	リスク比	95%信頼区間
母数効果	逆分散法	0.703	0.565～0.875	0.878	0.806～0.957
	Mantel-Haenszel 法	0.697	0.562～0.865	0.860	0.786～0.941
	Peto 法	0.698	0.563～0.865	—	—
変量効果	DerSimonian-Laird 法	0.670	0.517～0.870	0.868	0.779～0.967
	ベイズ流モデル	0.645	0.449～0.868	0.851	0.692～0.946

異質性を説明する要因の探索

- メタアナリシスを行う目的の一つは、個々の試験結果が食い違う原因の追究、すなわち**異質性を説明する要因の探索**である
- 異質性が小さい場合には、モデルによってメタアナリシスの結果は大きく異ならず、結果の解釈は比較的容易であるが、異質性が大きい場合には、その原因を探索することが重要である
- そのような探索的な解析によって、結果の解釈に新たな示唆を得ることができるし、新たな仮説が生まれる可能性もある
- 異質性の源となり得る要因は、以下のように分類できる
 - 個人レベルの特性：年齢、疾患重症度、合併症など
 - 試験レベルの特性：試験治療(薬剤、用量)など

サブグループ解析

- 異質性の原因を探索する最も単純な方法である
- 文献に基づくメタアナリシスでは、試験レベルの特性（試験治療の用量など）でサブグループを形成できる
- 個人データに基づくメタアナリシスでは、試験レベルの特性に加えて、様々な個人レベルの特性で集団を分類して治療効果を検討できる
- サブグループ解析には、多重性の問題などがあるが、メタアナリシスの場合にも同様の注意が必要である

メタ回帰

- 文献に基づくメタアナリシス -

- メタ回帰 (meta-regression) とは、各試験の効果の大きさの推定値を目的変数とし、試験レベルの特性を説明変数とした回帰分析である
- 試験 i の共変量の観測値を $x_{i1}, x_{i2}, \dots, x_{ip}$ としたとき、効果の大きさ θ_i に対する母数効果回帰モデルは、

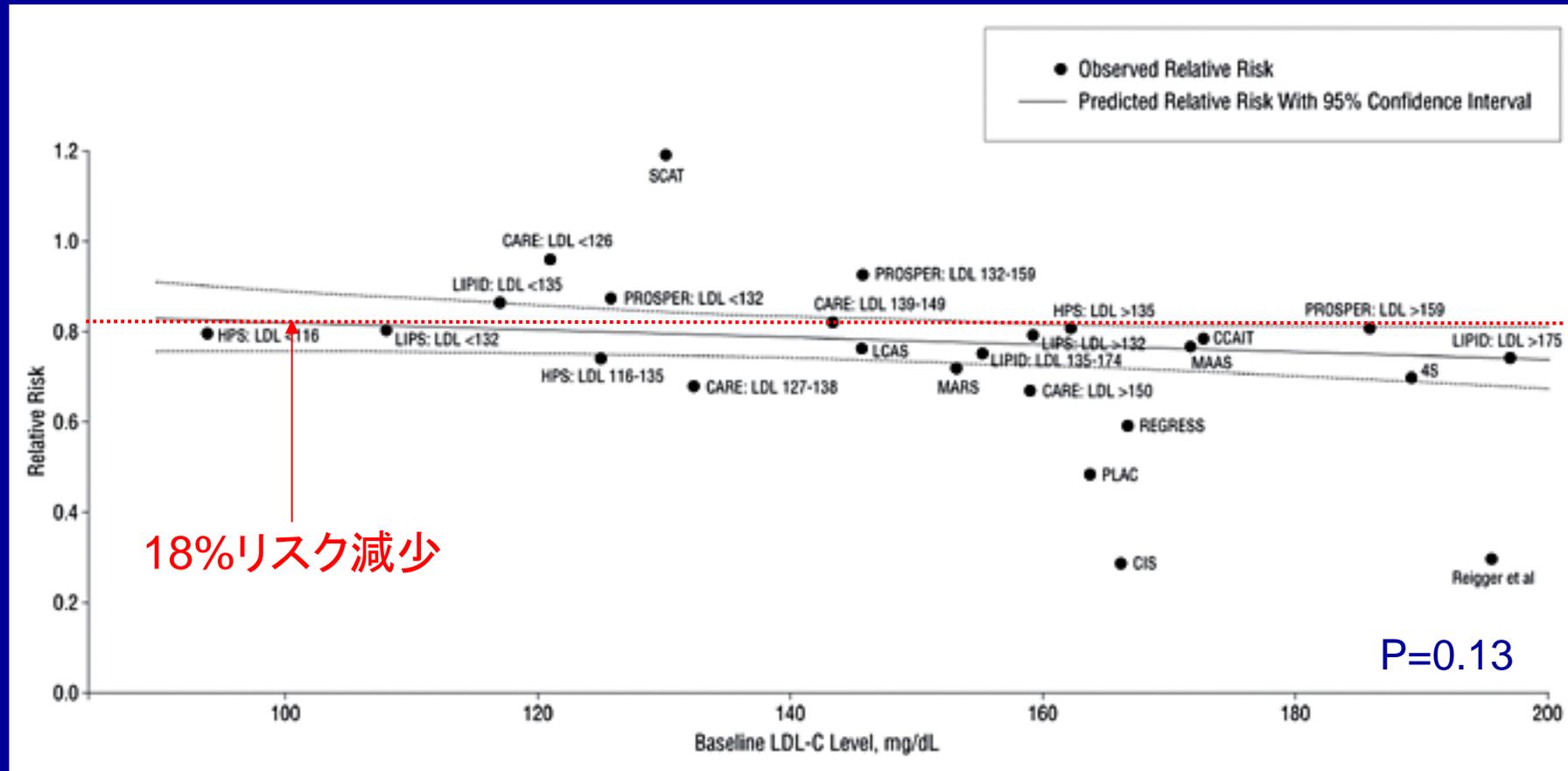
$$\theta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \quad (44)$$

ここで、 $\beta_0, \beta_1, \dots, \beta_p$ は未知の回帰係数である

- 各回帰係数の最尤推定量は、式(13)の w_i を重みとした重み付き最小二乗法により求めることができる

メタ回帰の例

- ベースラインLDL-Cとリスク比(主要血管イベント)の関係 -



Wilt TJ, et al. Arch Intern Med 2004;164:1427-1436

公表バイアス

- メタアナリシスから妥当な結論を得るために最も重要なことは、テーマに関連する試験が全て同定・評価されることである
- メタアナリシスにおける選択バイアスの問題は、ポジティブな試験が、ネガティブな試験に比べて公表されやすいという公表バイアス (publication bias) に関連して議論されることが多い
- 公表バイアスの存在は、1956年にある雑誌の編集者によって指摘されて以来、現在では広く認められている
- 公表バイアスは、引き出しの中に公表されない研究結果が眠っているという意味で、書類引き出し問題 (file-drawer problem) と呼ばれることもある

公表バイアスの検出

- 公表バイアスの検出には、小規模な試験の結果はバラツキが大きく、試験が大規模になるに従ってそのバラツキが小さくなることを利用した漏斗プロット (funnel plot) がよく用いられる

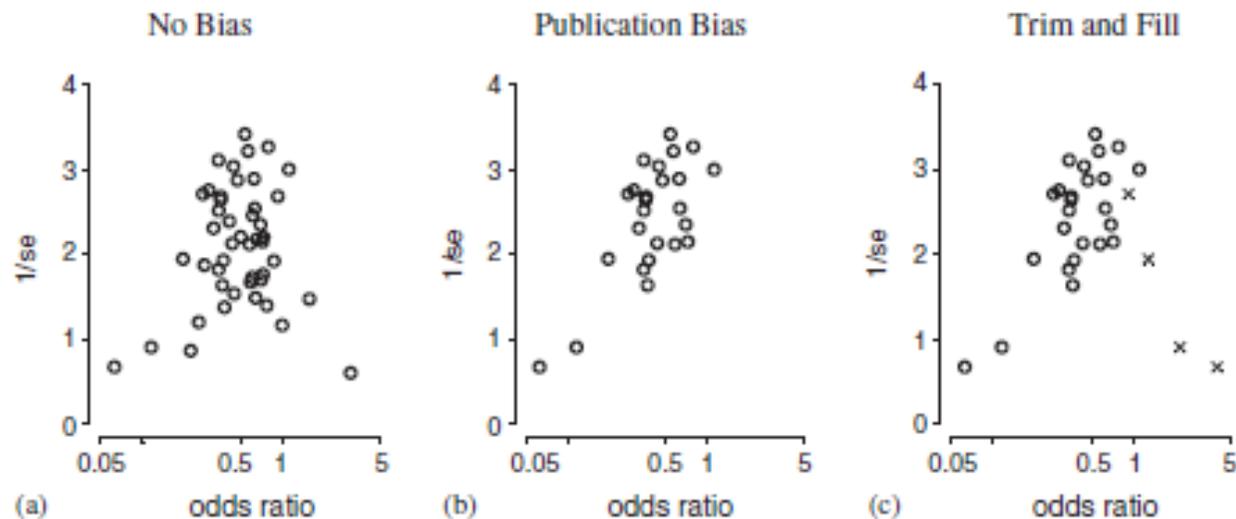


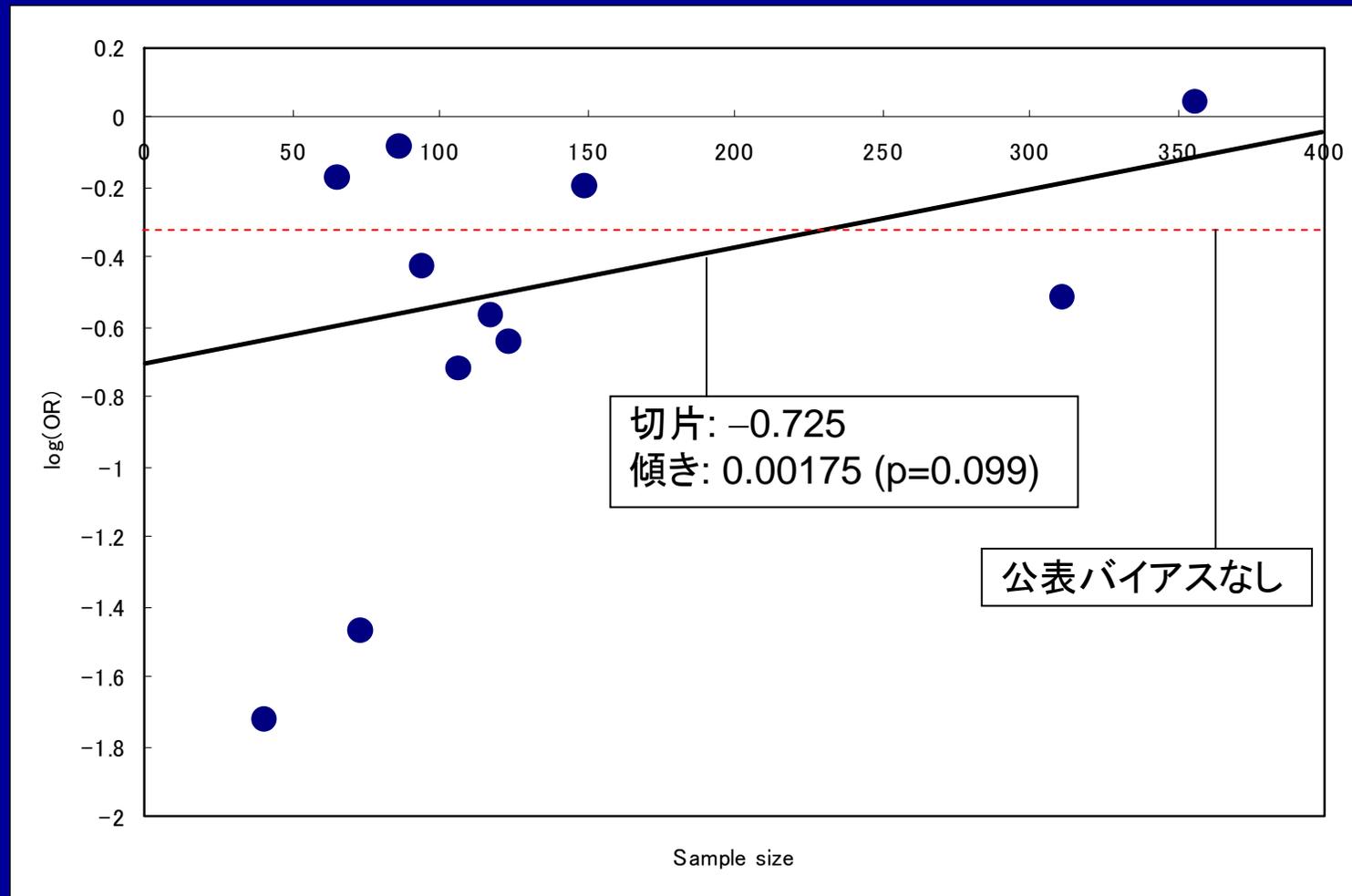
Figure 1. (a) Funnel plot of a simulated collection of trials with sample size 50–500 and true odds ratio 0.50. (b) A subset of the previous studies, with selection based on p -value and sample size. (c) Crosses were added to represent studies imputed by trim and fill.

- 漏斗プロットの一方の軸は効果の大きさの推定値であり、もう一方の軸は効果の大きさの推定精度(標準誤差の逆数、標本サイズなど)である
- 公表バイアスが小さければ、漏斗プロットが対称になるという性質を利用した公表バイアスの検出法として漏斗プロット回帰(Macaskill et al, 2001)がある
- これは、試験*i*の効果の大きさの推定値 T_i を目的変数、試験*i*の標本サイズ n_i を独立変数とした重み付き回帰分析である
- 重みとしては効果の大きさの分散の逆数を用いる

$$T_i = \alpha + \beta n_i + e_i$$

$$H_0 : \beta = 0$$

図3. 漏斗プロットと漏斗プロット回帰の例



公表バイアスの調整

- 公表バイアスの調整については、代表的なものは、選択モデルアプローチとTrim and Fill法である
- 選択モデルアプローチは、効果の大きさの推定値と公表される確率との関係(選択モデル)を特定して、統合推定値を調整する方法である
- Trim and Fill法は、「公表バイアスがなければ、漏斗プロットが対称となる」という前提の下に、隠されている可能性のある試験を補完して統合推定値を調整する方法である
- いかなる方法も検証できない仮定に基づいているので、調整というよりも感度分析に有用である

Trim and Fill 法

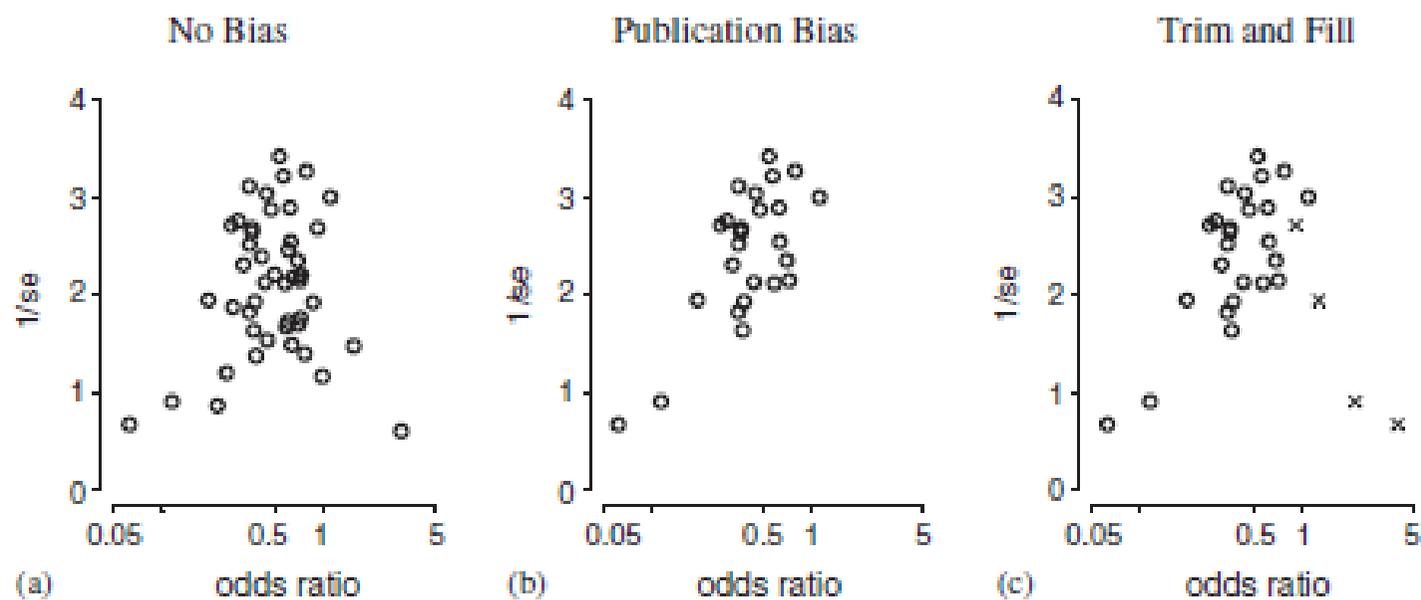


Figure 1. (a) Funnel plot of a simulated collection of trials with sample size 50–500 and true odds ratio 0.50. (b) A subset of the previous studies, with selection based on p -value and sample size. (c) Crosses were added to represent studies imputed by trim and fill.

公表バイアスの防止

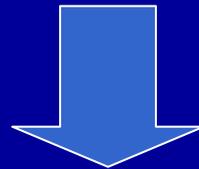
- 公表バイアスを防ぐ唯一の方法は、全ての臨床試験の登録と公表である
- 2004年、臨床試験の登録に関して、11のICMJE(医学雑誌編集者国際委員会)の会員雑誌で試験登録ポリシーが採択された
 - 雑誌が掲載を考慮する1つの条件として、公的な試験登録への登録を要求する
- 臨床試験の実施と報告に関する完全な透明性が最終目標であり、登録はその目標への第一歩である
- 今後、この流れはメタアナリシスの有用性を高め、その方法論を大きく変える可能性がある

メタアナリシスのためのソフトウェア

- 一般の統計解析ソフトウェア
 - SAS, Stata, S-Plus/R など
- ベイズ流解析
 - WinBUGS
- メタアナリシスに特化した商用ソフトウェア
 - Comprehensive Meta-Analysis, Stata macro など
- メタアナリシスに特化したフリーソフトウェア
 - RevMan, MIX (Excel add-in) など

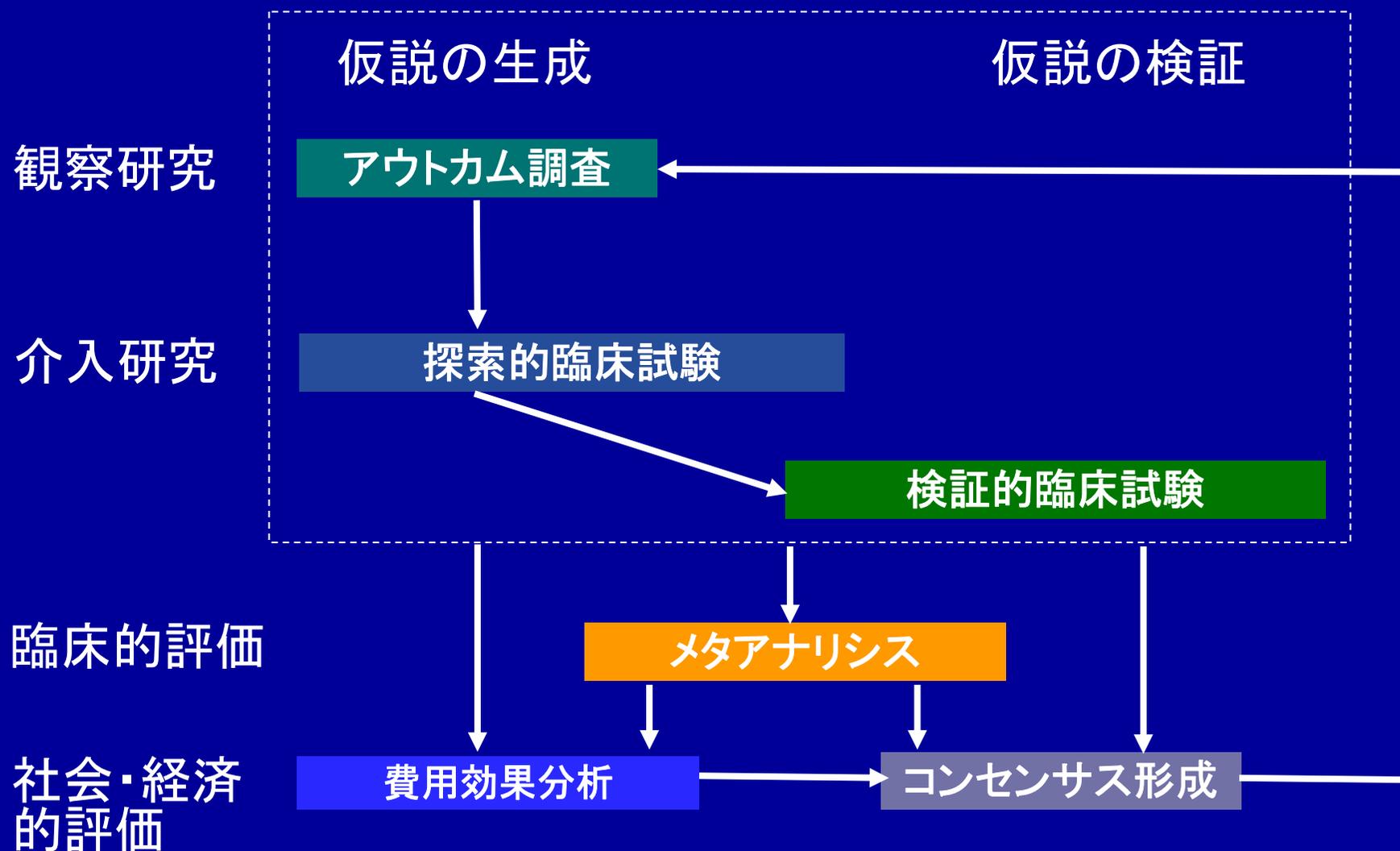
情報の統合

- すべての情報が利用可能ではない
- 個々の情報は完全ではない
- モデル化には多くの検証できない仮定が必要である



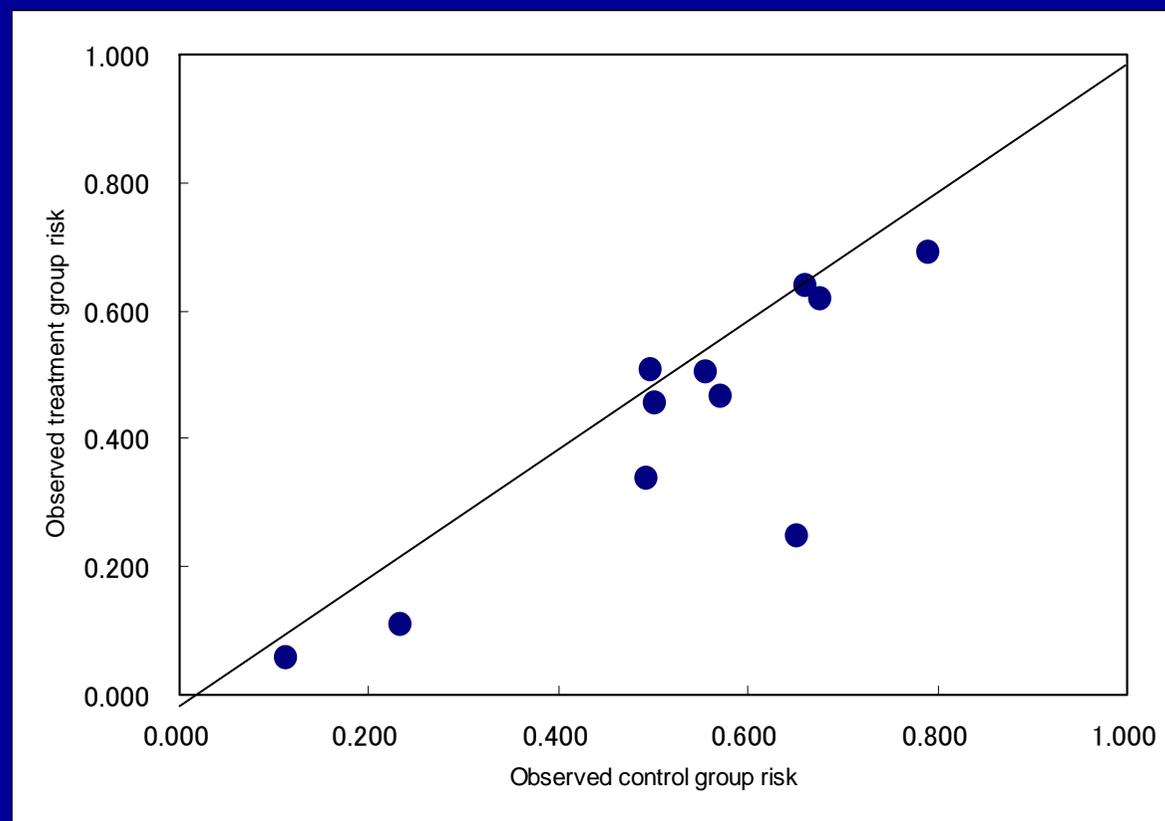
多くの仮定を置いて、不完全な情報を意思決定に役立つ形に統合するパッチワークの技術

医療技術評価のプロセス



Back-up

図2. L'Abbéプロットの例



- 結果が二値変数の場合には、各試験の治療群のイベントリスク(イベント数/対象数)を対照群のそれに対してプロットするL'Abbéプロットが有用である
- リスク比が試験間で均質であれば、点は原点を通るある直線の近くに集まり、そうでなければ、異質性が疑われる

- 効果の尺度がオッズ比の場合、

$$T_i = \log(OR_i) \quad (16)$$

$$v_i = \frac{1}{a_i} + \frac{1}{b_i} + \frac{1}{c_i} + \frac{1}{d_i} \quad (17)$$

- リスク比の場合、

$$T_i = \log(RR_i) \quad (18)$$

$$v_i = \frac{1}{a_i} - \frac{1}{n_{1i}} + \frac{1}{c_i} - \frac{1}{n_{0i}} = \frac{b_i}{a_i n_{1i}} + \frac{d_i}{c_i n_{0i}} \quad (19)$$

- リスク差の場合、

$$T_i = RD_i \quad (20)$$

$$v_i = \frac{a_i b_i}{n_{1i}^3} + \frac{c_i d_i}{n_{0i}^3} \quad (21)$$

を代入すれば、それぞれの統合推定値が得られる

- 効果の尺度が平均値あるいは標準化された平均値の場合にも、同様にして統合推定値を得ることができる

Mantel-Haenszel法

- 交絡因子を調整するための方法としてよく知られており、各試験の対象数が少ないときには他の方法に比べて性質が良い
- 効果の尺度がオッズ比の場合、重み w_i を用いて、式(12)から、統合オッズ比を得る

$$w_i = \frac{b_i c_i}{n_i} \quad (22)$$

$$OR = \frac{\sum_{i=1}^k a_i d_i / n_i}{\sum_{i=1}^k b_i c_i / n_i} \quad (23)$$

- 統合オッズ比の分散の推定にはいくつかの方法が提案されている
- 効果の尺度がリスク比、リスク差、ハザード比の場合にも利用できる

Peto法

- 効果の尺度としてオッズ比を用いる場合に考案された方法で、Mantel-Haenszel変法と呼ばれる
- 試験*i*の治療群における観測イベント数 O_i (表1では a_i)と治療効果がない場合の期待イベント数 E_i (表1では $n_{1i}m_{1i}/n_i$)との差およびその分散 V_i から近似的に推定されるオッズ比

$$OR_i = \exp\left(\frac{O_i - E_i}{V_i}\right) \quad (24)$$

を用いる。ここで、

$$V_i = \frac{n_{1i}n_{0i}m_{1i}m_{0i}}{n_i^2(n_i - 1)} \quad (25)$$

- 統合オッズ比は、

$$OR = \exp\left(\frac{\sum_{i=1}^k (O_i - E_i)}{\sum_{i=1}^k V_i}\right) \quad (26)$$

と推定され、その95%信頼区間は

$$\exp\left(\frac{\sum_{i=1}^k (O_i - E_i)}{\sum_{i=1}^k V_i} \pm 1.96 \sqrt{\frac{1}{\sum_{i=1}^k V_i}}\right) \quad (27)$$

- セルにゼロがある場合でも計算可能という利点はあるが、各試験のオッズ比が1から離れると、重大な過小推定になること、群間の対象数が大きく異なると重大なバイアスが生じることなどが知られている

DerSimonian-Laird改良法

- 式(30)、(31)の w_i に $1/v_i$ を用いて、

$$\hat{\tau}^2(DL) = \max\left\{0, \frac{Q - (k - 1)}{U}\right\}$$

を推定した後、

$$w_i = \frac{1}{\hat{\tau}^2(DL) + v_i}$$

を用いて、同じ手順で τ^2 のモーメント推定値を得る(2段階モーメント法)

Paule-Mandel法

- 以下の推定方程式を反復計算により解き、 τ^2 を推定する

$$F(\tau^2) = \sum_{i=1}^k \frac{(T_i - \bar{T})^2}{W_i} - (k-1) = 0$$

ここで、

$$W_i = \frac{1}{\tau^2 + v_i}$$

$$\bar{T} = \frac{\sum_{i=1}^k W_i T_i}{\sum_{i=1}^k W_i}$$

- 正規性の仮定の下で制限付き最尤(REML)推定値となり、統合推定値も統計的最適性を有する

ベイズ流モデル

- 式(28)の変量効果モデルでは、

$$T_i \sim N(\theta_i, v_i) \quad (38)$$

$$\theta_i \sim N(\mu, \tau^2) \quad (39)$$

と仮定し、確率分布のパラメータ(超パラメータ) μ と τ^2 を定数と考えている

- 超パラメータに確率分布を考えるアプローチをベイズ流アプローチ(Bayesian approach)と呼ぶ
- このようなモデルは、ベイズ流階層モデル(hierarchical model)とも呼ばれる
- ベイズ流アプローチでは、全てのパラメータの不確実性を考慮でき、柔軟性と拡張性に優れている

- 効果の尺度がオッズ比の場合、次のようなモデル化が可能である

ここで、パラメータ π_{0i} 、 π_{1i} はそれぞれ対照群、治療群の真のイベント発生割合、 μ_i は真の潜在リスク(対照群のイベント発生割合に関するロジット)、 δ_i は真の治療効果を表す

$$\begin{aligned}c_i &\sim \text{Binomial}(\pi_{0i}, n_{0i}) \\a_i &\sim \text{Binomial}(\pi_{1i}, n_{1i}) \\ \mu_i &= \text{logit}(\pi_{0i}) = \log\left(\frac{\pi_{0i}}{1 - \pi_{0i}}\right) \\ \text{logit}(\pi_{1i}) &= \mu_i + \delta_i\end{aligned}\tag{42}$$

$$\delta_i \sim N(\delta, \tau^2)$$

$$\delta \sim [-, -]$$

$$\tau \sim [-, -]$$

SAS Code

- Mantel-Haenszel法 & 逆分散法 -

```
data meta;
  input trial trt event count @@;
datalines;
1 1 1 110 1 1 2 49
1 2 1 120 1 2 2 32
...
;
proc freq data=meta;
  weight count;
  tables trial * trt * event / relrisk cmh;
run;
```

Trial 1	Event	No Event
Treatment	110	49
Control	120	32



試験1の結果

表 1 : trt * event

層別変数 : trial=1

trt	event		合計
	1	2	
1	110 69.18	49 30.82	159
2	120 78.95	32 21.05	152
合計	230 73.95	81 26.05	311

相対リスクの推定値 (行 1 / 行 2)

研究の種類	値	95% 信頼区間	
ケースコントロール研究 (オッズ比)	0.5986	0.3576	1.0021
コホート研究 (列 1 の相対リスク)	0.8763	0.7677	1.0003
コホート研究 (列 2 の相対リスク)	1.4638	0.9951	2.1535

全体の結果

Cochran-Mantel-Haenszel 統計量(テーブルスコアに基づく)

統計量	対立仮説	自由度	値	p 値
1	相関統計量	1	10.7829	0.0010
2	ANOVA 統計量	1	10.7829	0.0010
3	一般関連統計量	1	10.7829	0.0010

相対リスクの推定値 (行 1 / 行 2)

研究の種類	調整方法	値	95% 信頼区間	
ケースコントロール研究	Mantel-Haenszel	0.6970	0.5616	0.8650
(オッズ比)	ロジット	0.7028	0.5648	0.8745
コホート研究	Mantel-Haenszel	0.8600	0.7858	0.9413
(列 1 のリスク)	ロジット	0.8782	0.8062	0.9567
コホート研究	Mantel-Haenszel	1.1817	1.0690	1.3063
(列 2 のリスク)	ロジット	1.1325	1.0455	1.2267

ロジット=
逆分散法

オッズ比等質性に対する
Breslow-Day 検定

カイ 2 乗値	13.0136
自由度	10
Pr > ChiSq	0.2229

解析手法

- 個人データに基づくメタアナリシス -

結果の異質性

- 個人データが利用できる場合、結果の異質性の検定は、回帰モデルにおける試験と治療効果との交互作用の検定に帰着し、これは式(8)のQ統計量に基づく検定に対応する
- 一般に、効果の尺度が平均値の差の場合は一般線形回帰モデル、オッズ比の場合はロジスティック回帰モデル、生存期間の場合には比例ハザード回帰モデルを用いる

結果の要約

- メタアナリシスにおける試験を、多施設試験における施設と同様に扱えば、個人データに基づくメタアナリシスと多施設試験データの解析とは多くの共通点を持つ
- 文献に基づく場合に比べて、個人データを用いれば様々なモデル化が可能である
- 共変量データが収集されていれば、それらについて調整した統合推定値を得ることができる

母数効果モデル

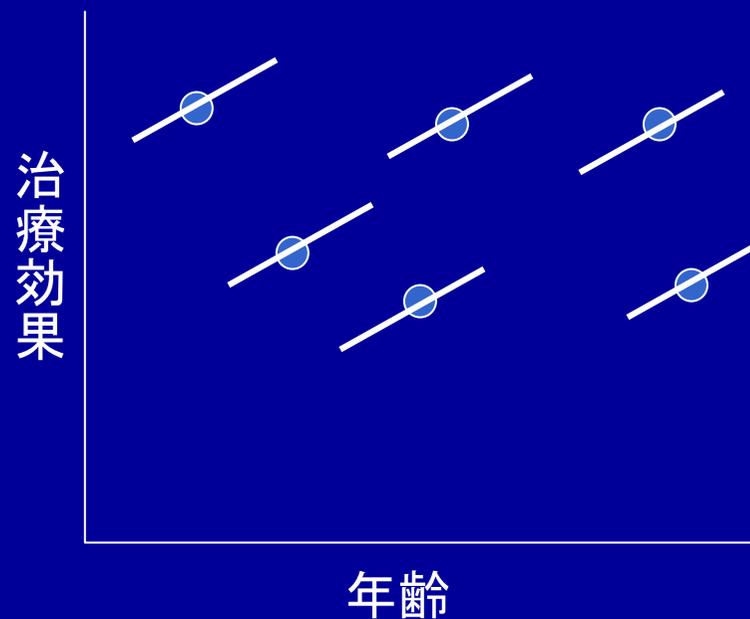
- 各試験の真の効果の大きさは同一と仮定した上で、回帰モデルの群に関するパラメータ推定値から統合推定値を得ることができる
- 回帰モデルにおいて、「試験」は特別な変数として扱う必要があり、例えば生存時間がエンドポイントの場合には、試験ごとに対照群リスク(基準ハザード)が異なることを許容した、層別比例ハザードモデルを用いるのが一般的である

変量効果モデル

- 様々なモデル化が可能である
- 効果の尺度が平均値の差の場合は一般線形混合モデル、二値変数とときは一般化線形混合モデル(ロジスティックモデルやポワソンモデルの拡張)、生存期間のときには比例ハザードモデルを拡張したモデルが利用できる

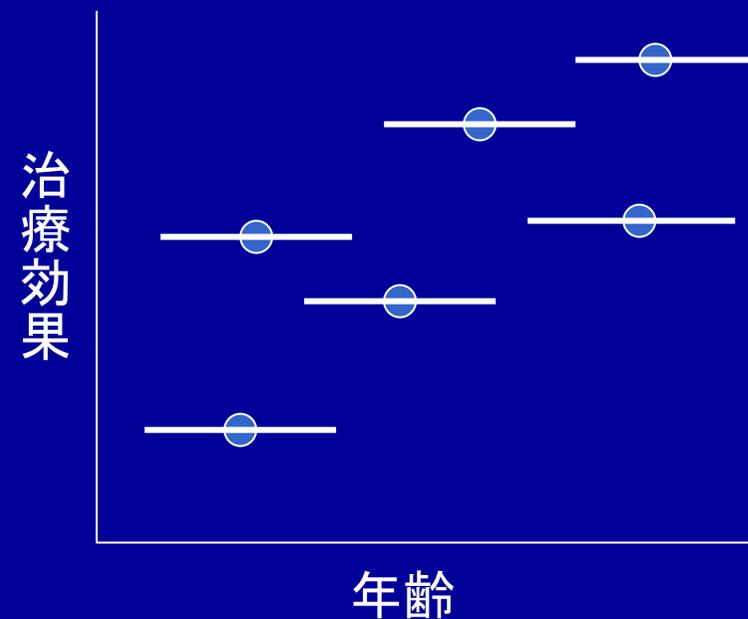
- 式(44)の母数効果回帰モデルに変量を導入して混合効果モデルに拡張することも、ベイズ流の解析を行うことも可能である
- 対照群における死亡割合などの対照群リスクを共変量として治療効果との関連性を調べる手法も数多く開発されている
- しかしながら、試験レベルでの要約値を含んだモデルについては、集団レベルの共変量と治療効果との関連が必ずしも個人レベルの共変量と治療効果を反映していないという集団バイアス (aggregation biasあるいは ecological bias) が指摘されている

集団バイアス



試験内の関連: ○

平均値と治療効果の関連: ×



試験内の関連: ×

平均値と治療効果の関連: ○

回帰分析

- 個人データに基づくメタアナリシス -

- 個人レベルで測定された共変量を用いれば、メタ回帰の問題の多くを克服でき、より信頼性の高い解析を行うことが可能である
- 異質性の原因を探索するという目的においては、様々なモデルを構築でき、ある個人レベルの変数と治療効果との関連に関心がある場合には、それらの変数と治療効果との交互作用項を回帰モデルに入れて評価できる

拡張・他の応用領域

- 拡張
 - － 多重パラメータ(複数エンドポイント)
 - － 異なるデザインの研究(ランダム化試験と観察研究)の統合
 - － 混合治療法比較(直接比較、間接比較の試験を統合)
- 他の応用領域
 - － 費用-効果分析
 - － 代理エンドポイントの評価
 - － 遺伝子相関研究
 - － 診断・スクリーニング研究
 - － 予後因子研究