



# メタアナリシス入門

横浜市立大学大学院医学研究科 臨床統計学

山中 竹春

# 本日本話しする内容

- **メタ解析の定義、フォレストプロットの見方**
- **ハザード比を統合するときの考え方**
- **RCTのメタ解析の主な役割**
- **試験間に結果の異質性が見られる場合は？**
- **固定効果モデル vs. 変量効果モデル？**
- **まとめ**



# メタ解析の定義、 フォレストプロットの見方



# メタ解析 (Meta-Analysis) とは

---

- 教育学者 Glass による造語

**Meta-** 超えて (beyond), 高次の (higher),  
後から (after) などを意味する接頭語

Glass GV, *Educational Res* 5, pp.3-8, 1976

- 独立に実施された複数の「同種」研究  
を統計的に統合したもの

# 系統的レビュー (Systematic Review) との関係



## Systematic review

- A review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research, and to collect and analyze data from the studies that are included in the review. Statistical methods (**meta-analysis**) may or may not be used to analyze and summarize the results of the included studies.

## Meta-analysis

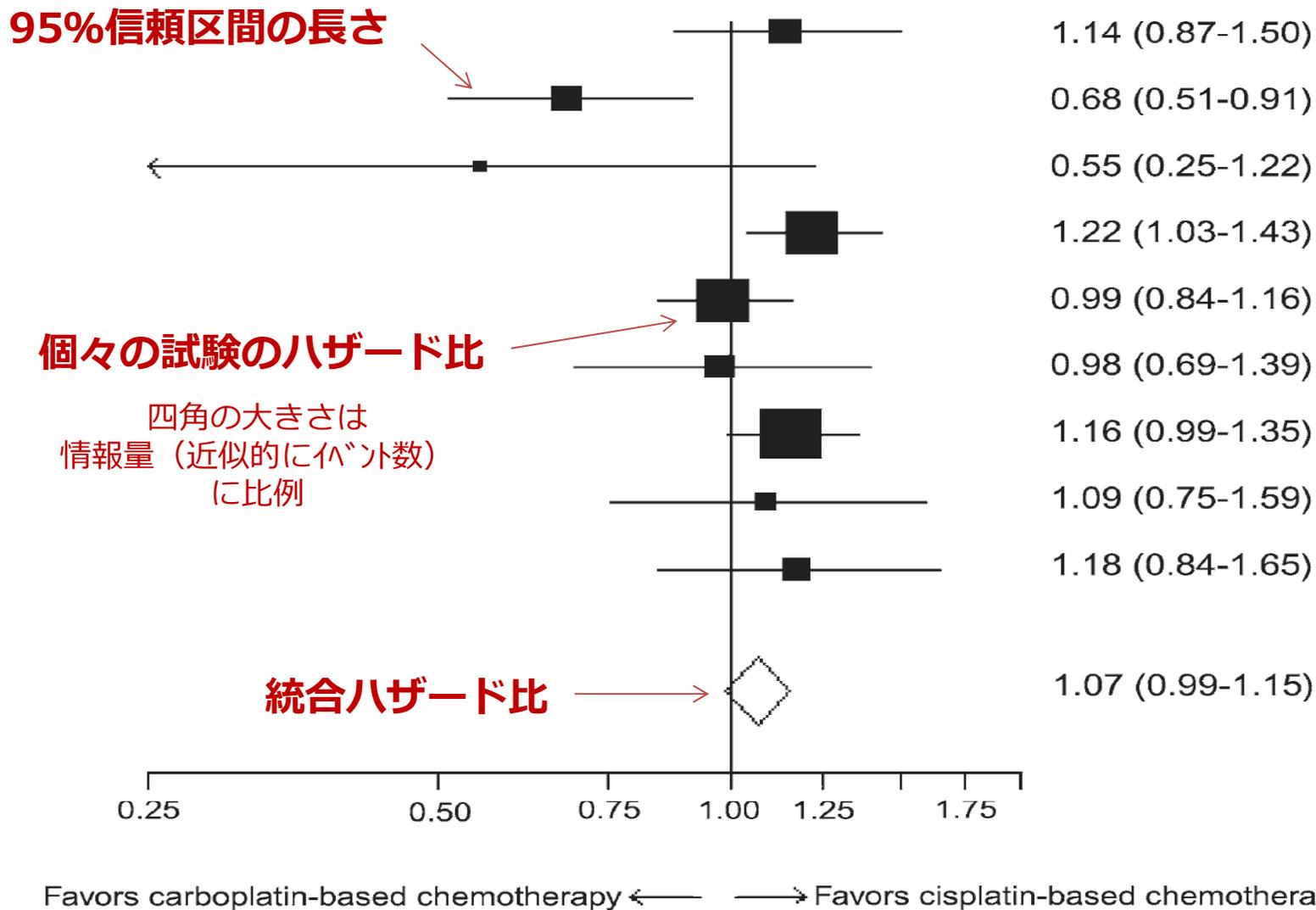
- The use of statistical techniques in a systematic review to integrate the results of included studies. Sometimes misused as a synonym for systematic reviews, where the review includes a meta-analysis. **(系統的レビューにおける統計解析部分に相当)**

“Glossary” in Cochran Library  
<http://community.cochrane.org/glossary/5#letterm>

# 事例：進行期NSCLC CBDCA-based vs. CDDP-based



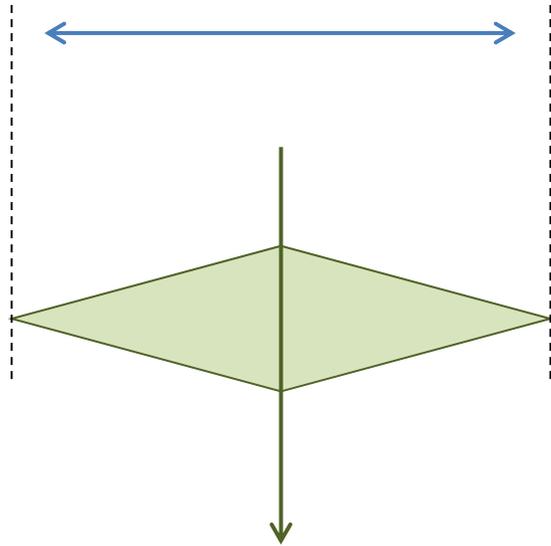
Ardizzone *et al*, JNCI 99, p847-57, 2007



# 統合ハザードの見方

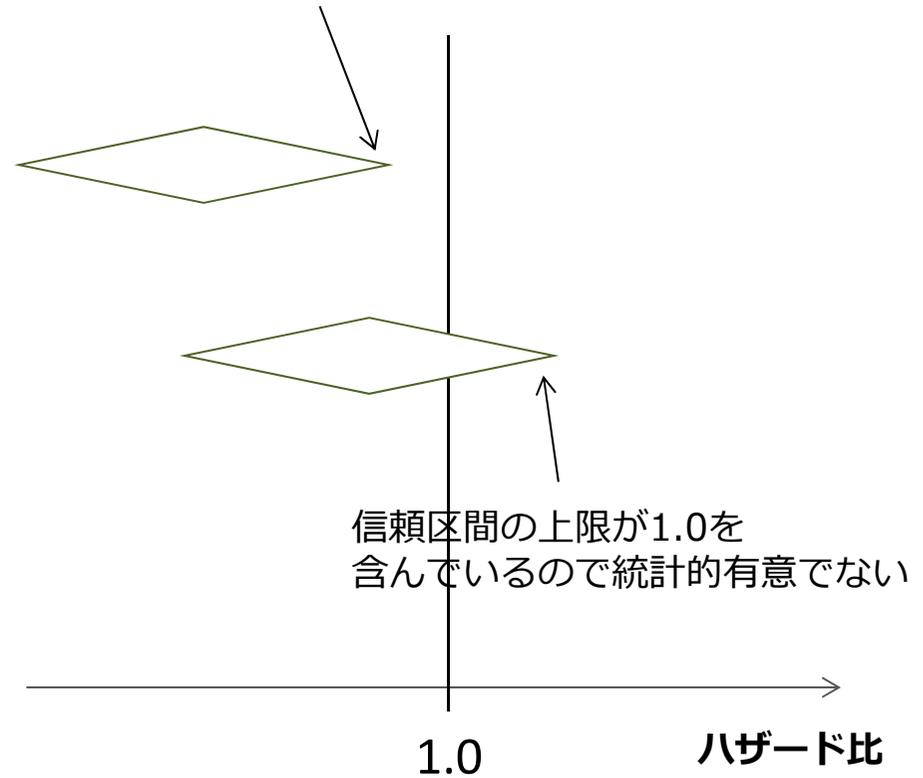


統合ハザード比の95%信頼区間



統合ハザード比の値

信頼区間の上限が1.0を  
含んでいないので統計的有意





# ハザード比（HR）を 統合するときの考え方



# メタ解析で仮定される統計モデル

- **固定効果モデル** (fixed-effects model)
  - 真のHRがひとつ存在して、試験間のHRのバラツキは、その値からの「偶然誤差」である  
と考える

# 固定効果モデルのイメージ



**小規模試験**  
イベント数：小

観測値  
(HR=0.62)

偶然  
誤差

**中規模試験**  
イベント数：中

観測値  
(HR=0.77)

偶然  
誤差

**大規模試験**  
イベント数：大

観測値  
(HR=0.92)

偶然  
誤差

観測される効果



各試験共通の真の効果  
+  
偶然誤差

各試験共通の真の効果

Log (HR)

# 固定効果モデルにおける 統合方法の例（逆分散法）

各試験の効果の推定値  $T_1, T_2, T_3$  をそれらの分散の逆数  $w_i, i=1, 2, 3,$  で重み付け平均をとる方法

$$\text{統合値} = \frac{w_1 T_1 + w_2 T_2 + w_3 T_3}{w_1 + w_2 + w_3}$$

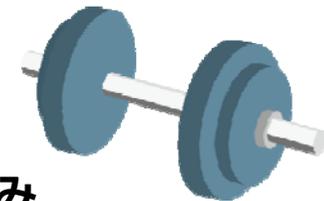
小さい重み  
イベント数小



中くらいの重み  
イベント数中



大きな重み  
イベント数大



すべての試験でイベント数が同程度なら、単純平均に近くなる



# メタ解析における統合方法の選択

---

- 逆分散法
  - Petoの方法
  - Mantel-Haenszelの方法
- etc.

**原理（考え方）はいずれも重み付け平均  
どの手法を使っても大きく結果が異なる  
わけではない**



# RCTのメタ解析の主な役割

## RCTのメタ解析の主な役割

- 単一研究のみでは見出しにくい、小さな、しかし臨床的に意味のある差に対して、**精度の高い推定**を行える
  - 検定よりも、むしろ推定
- 症例数の点で検討が難しい**サブグループ****に対する検討**が可能
- 個々の試験の結果を踏まえ、より一般的な結果を導けるか（**一般化可能性**）、試験間に異質性が見られないか、を検討



## RCTのメタ解析の主な役割

---

- 単一研究のみでは見出しにくい、小さな、しかし臨床的に意味のある差に対して、**精度の高い推定**を行える
  - 検定よりも、むしろ推定
- 症例数の点で検討が難しいサブグループに対する検討が可能
- 個々の試験の結果を踏まえ、より一般的な結果を導けるか（一般化可能性）、試験間に異質性が見られないか、を検討



# RCTのメタ解析の主な役割

---

- 単一研究のみでは見出しにくい、小さな、しかし臨床的に意味のある差に対して、精度の高い推定を行える
  - 検定よりも、むしろ推定
- 症例数の点で検討が難しい**サブグループ**  
**に対する検討**が可能
- 個々の試験の結果を踏まえ、より一般的な結果を導けるか（一般化可能性）、試験間に異質性が見られないか、を検討



# RCTのメタ解析の主な役割

- 単一研究のみでは見出しにくい、小さな、しかし臨床的に意味のある差に対して、精度の高い推定を行える
  - 検定よりも、むしろ推定
- 症例数の点で検討が難しいサブグループに対する検討が可能
- 個々の試験の結果を踏まえ、より一般的な結果を導けるか（**一般化可能性**）、試験間に異質性が見られないか、を検討



# 試験間に結果の異質性が 見られる場合は？ ～異質性を評価する手法～



- 試験間に、“偶然誤差”, “ランダムな誤差”を超えた結果のバラつき（異質性）が認められる場合、どうするか？
  - 単純に、イベント数に応じた重み付き平均（固定効果モデルによる）でガッチャンコしてよいか？ という問題



# 「結果の異質性」を評価する統計手法

## 異質性の検定 (Test of Heterogeneity)

- コクランのQ検定
- $P$ 値が小さければ「試験結果に異質性がある」
- 試験数が少ないときは検出力は高くない。 $P < 0.10$ が規準

## $I^2$ index

- 0~100%の値をとる。試験結果の不均一性の尺度
- (目安)  $I^2 \approx 25\%$ で小さな不均一性、 $I^2 \approx 50\%$ で中程度の不均一性、 $I^2 > 75\%$ で大きな不均一性

# 「結果の異質性」を評価する統計手法

$$Q = \sum_{i=1}^K w_i (d_i - d_s)^2 \sim \chi_{K-1}^2$$

$K$  試験の数

$w_i$  試験  $i$ ,  $i = 1, 2, \dots, K$ , の重み

$d_i$  試験  $i$ ,  $i = 1, 2, \dots, K$ , の群間差

$d_s$   $K$  個の試験の群間差の統合値

$$I^2 = 100\% \times (Q - (K - 1)) / Q$$

# 「結果の異質性」を評価する統計手法

Ardizzoni *et al*, JNCI 99, p847-57, 2007

## Statistical Analysis

The primary endpoint in the meta-analysis was overall survival, defined as the time elapsing from random assignment until death from any cause. Living patients were censored at the date of last follow-up. The secondary endpoints were overall response rate, defined as the sum of partial and complete response rates (according to World Health Organization criteria), and toxicity, which was graded according to NCI Common Toxicity Criteria (CTC).

All analyses were conducted on an intention-to-treat basis, and all randomly assigned patients were included in the analyses according to the allocated treatment. A general variance-based method was used to estimate the summary hazard ratios (HRs), odds ratios (ORs), and their 95% confidence intervals (CIs) (16,17). Ratios were uniformly expressed relative to patients who received cisplatin therapy. The global null hypothesis that the treatment difference in all studies is equal to 0 was tested by comparing the  $U$  statistic with the chi-square distribution with 1 df (17). Cochran's chi-square test ( $Q$  test) was used to test for the presence of heterogeneity among trials (18). Moreover, the  $I^2$  index, which expresses as a percentage the proportion of variability of the results due to heterogeneity as opposed to sampling error, was calculated (19). Based on the statistical significance of the  $Q$  test, we applied a random-effects model, which allows the meta-analysis to take into account interstudy variation. In particular, the random-effects sur-

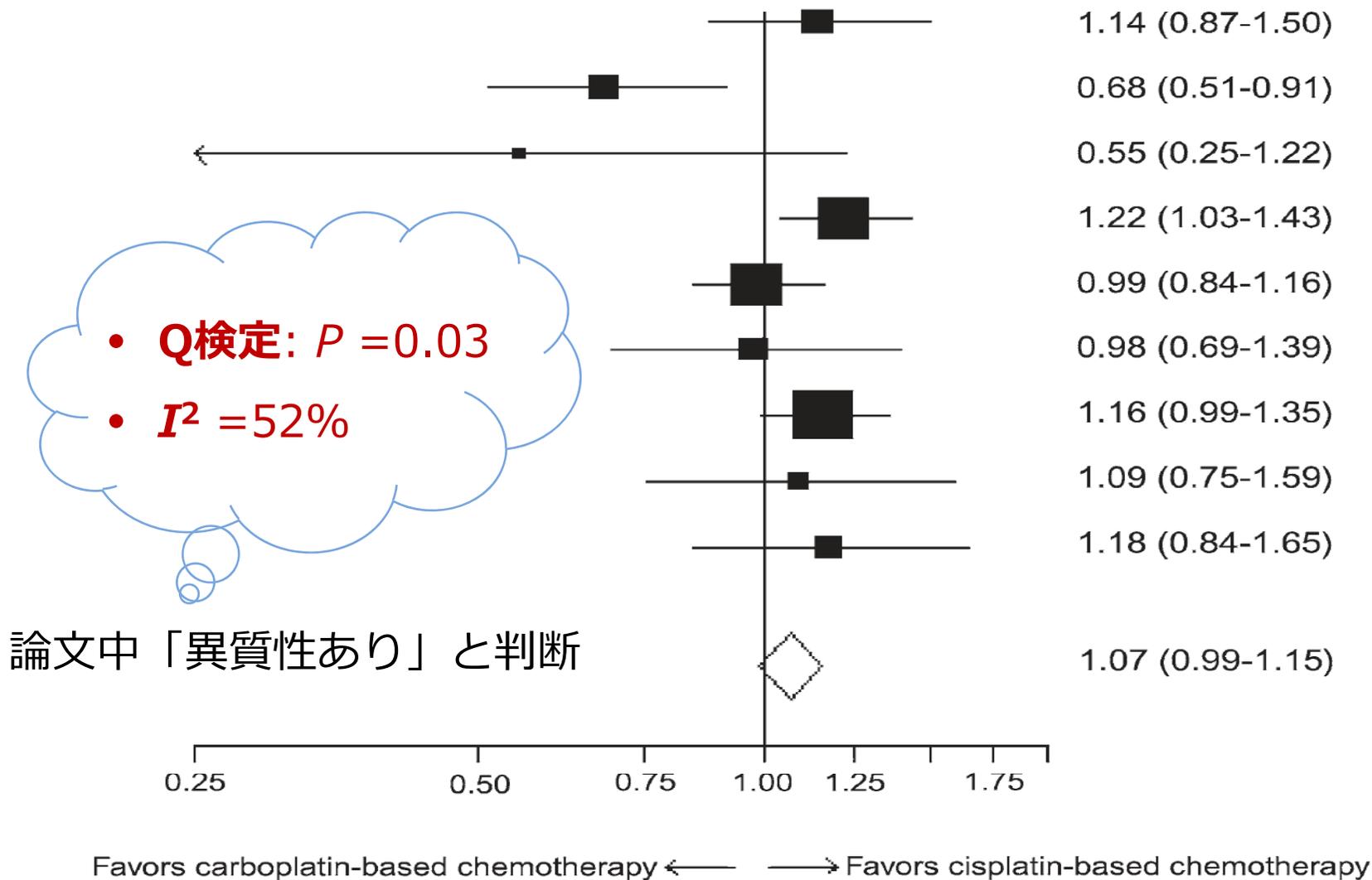
コクランのQ検定により試験間の異質性の存在について検定した

試験結果のばらつきの中で「試験間の異質性」により説明される割合%である $I^2$ を計算した

# 事例：進行期NSCLC CBDCA-based vs. CDDP-based



Ardizzoni *et al*, JNCI 99, p847-57, 2007





# 試験間に結果の異質性が 見られる場合は？ ～異質性を起こす要因の検討～



# 試験間の異質性が大きい場合は その原因となる要因の探索を !!

- **異質性を起こしがちな要因**

- 介入方法（薬剤，用量，投与スケジュール）、試験実施地域、年代の違いなど
- 年齢，重症度，性別など

- **感度分析による検討**

- どの部分集団で結果が変化するかを検討するサブグループ解析の実施

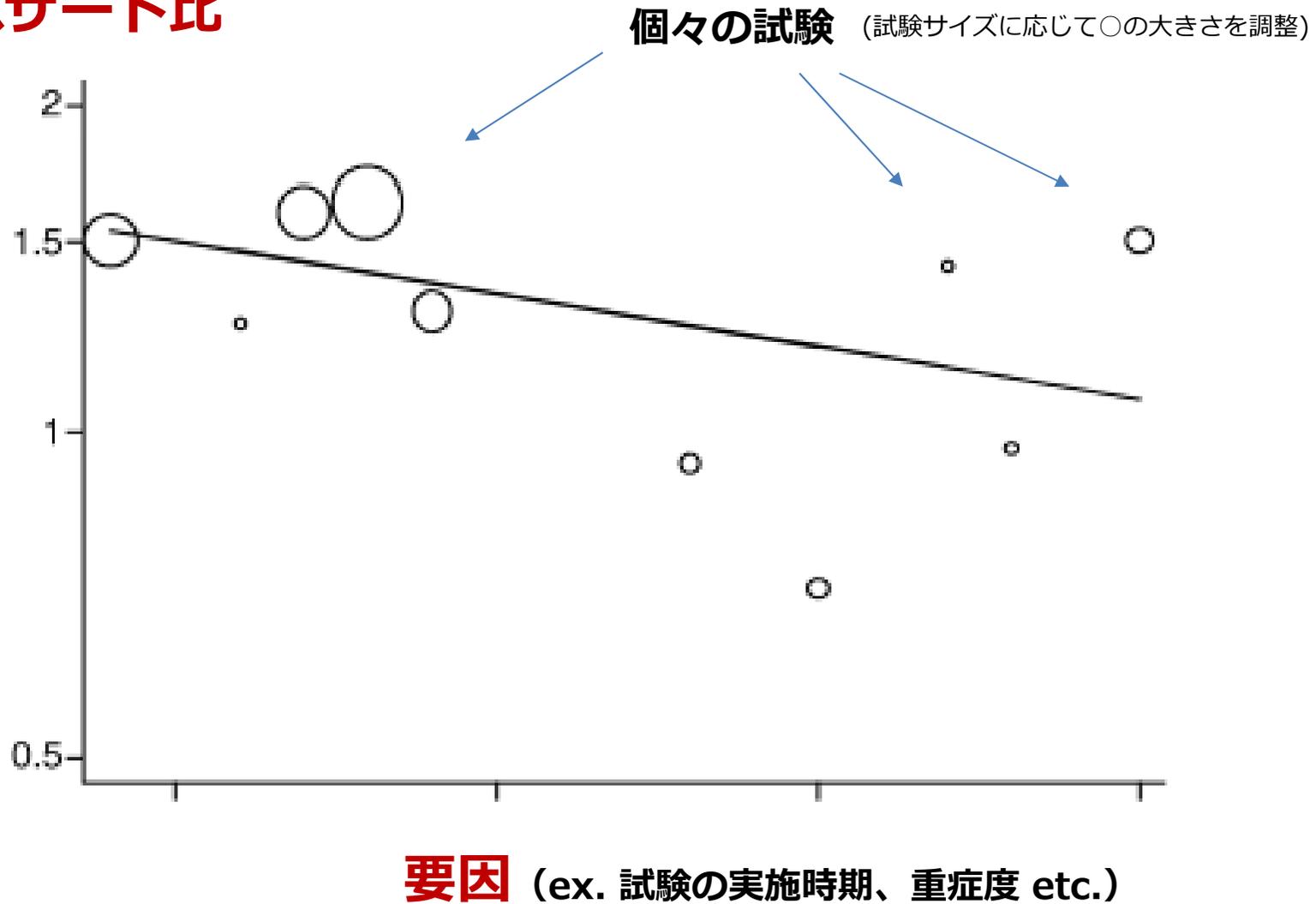
- **メタ回帰による検討**

- 要因を横軸に、効果を縦軸にとって回帰分析

# メタ回帰のイメージ



## ハザード比





## 補足：個別データか？出版データか？

- **Individual Patient Data (IPD) メタ解析**
  - 様々なサブグループ解析が可能
  - 複数因子の組み合わせまで含めた解析が可能  
(治療×因子の交互作用)
  - 多数の因子を調整した多変量解析が可能
  - 場合によっては、欠測や項目間の整合性を調べることで研究の質を評価

など



# 試験間に結果の異質性が 見られる場合は？ ～モデル化による対処～

# 固定効果モデルのイメージ



小規模試験  
イベント数：小

観測値  
(HR=0.62)  
偶然誤差

中規模試験  
イベント数：中

観測値  
(HR=0.77)  
偶然誤差

大規模試験  
イベント数：大

観測値  
(HR=0.92)  
偶然誤差

観測される効果



各試験共通の真の効果  
+  
偶然誤差

各試験共通の真の効果

Log (HR)

# 変量効果モデルのイメージ



**小規模試験**  
イベント数：小

観測値  
(HR=0.62)

← 偶然誤差

**中規模試験**  
イベント数：中

観測値  
(HR=0.77)

← 偶然誤差

**大規模試験**  
イベント数：大

観測値  
(HR=0.92)

← 偶然誤差

**試験間で真の効果が  
バラつくことを許容  
してしまう**

真の効果の"分布"

Log (HR) 30



# 小括：基本的な統計モデル

---

## 1. 試験内のバラつき

- 共通した真のハザード比があり、試験サイズ (イベント数) に応じたランダムな誤差を想定

## 2. 試験間のバラつき

- 真のハザード比自体が、一定の分布に従ってバラつくことを許容 (仮定)

**固定効果モデルでは1だけを仮定**

**1だけではバラつきを説明しきれない場合、  
2も仮定するのが変量効果モデル**

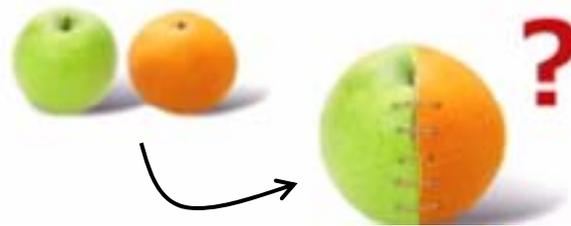




# 固定効果モデル vs. 変量効果モデル？

# 固定効果モデル vs. 変量効果モデル

- 研究間の異質性が小さい場合は、どちらのモデルを用いたメタ解析でも似たような結果になる
- 研究間の異質性が大きい場合は、異質性を説明する原因の探索がまず重要
  - 研究間の効果に異質性がある場合は、直ちに一般化可能性 (generalizability) に言及できない





# 固定効果モデル vs. 変量効果モデル

---

- 『異質性の検定を行って「異質性あり」と判断された場合は、変量効果モデル』と述べている教科書もあるが、そんなに単純な問題ではない
  - 変量効果モデルと固定効果モデルの結果が大きく食い違うようであれば、そもそも統合自体に問題のある可能性もある



# まとめ

## メタ解析：留意点

- **解釈の曖昧さを引き起こす問題を内包**
  - “同種”の研究をどう定義するか
  - 異質性を説明する原因の探索に限界がある場合も
  - 出版バイアスをいかに避けるか
- **個々の研究の質が、メタ解析の質を決定**
  - Garbage In, Garbage Out (GIGO)

## まとめにかえて

### (FAQ) 一つの大きなランダム化試験の結果とメタ解析のどちらが重いか？

- 一般には、単独の大きなランダム化試験をメタ解析が置き換えられるわけではなく、その逆でもない
- 相互補完的な立場
- 「治療法の選択」の意思決定には反映しづらい場合も少なくない

## まとめにかえて

### (FAQ) メタ解析で $P < 0.05$ であれば、 レベル1Aのエビデンスか？

- “同種の試験”の定義が適切であり、
- 問題になるほどの異質性がみられず、
- 個々の試験の質が高く、
- 出版バイアスがなければ・・・

メタアナリシス論文の質はさまざま

**私見では “Two Pivotal Studies” がレベル1A**