



〈メタ解析 (Forest Plot など) の読み方〉

1) Forest Plot の解釈

(1) Forest Plot とは

Forest Plot は、下図のようにメタ解析の結果を図に表したもので、複数の研究結果と統合した結果を視覚的に確認することができます。例題として図 1 に、VI 章 CQ13 「予防的抗菌薬によって、抗菌薬なしに比べて致命率は改善するか」を用いて解説いたします。フォレストは森林という意味であり、一つひとつの研究報告は木に例えられています。

例題の図 1 をご覧ください。Study of Subgroup の 1 つ目の論文は、Pederzoli et al. の論文です。予防的抗菌薬を 41 例に投与し 3 例が死亡した群に対して、Control (抗菌薬非投与群) では、33 例中 4 例が死亡しています。論文の記載では、オッズ比で 0.57, 95%信頼区間が 0.12 から 2.76 と記載され、1993 年に出版されています。

では、オッズ比, 95%信頼区間, さらに、統合結果について説明します。

(2) オッズ比の考え方と 95%信頼区間

オッズ比は、ある事象 (「Events」: 例題では「死亡」) の起こりやすさを 2 群 (抗菌薬「Experimental」群対コントロール「Control」群) で比較して示す統計方法です。オッズ比が 1 の場合は、2 群における事象の起こりやすさが同じ事を示しています (図 1 の右の Forest Plot の縦線がオッズ比 1 を示します)。オッズ比が 1 より小さければ、介入群 (抗菌薬群) で事象 (死亡) が、起こりにくいという事を示します。

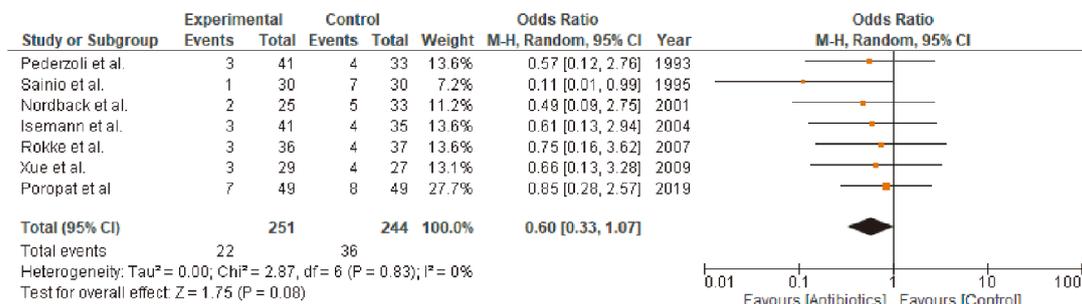


図 1 例題, 第 VI 章 CQ13 予防的抗菌薬の有無での致命率 (来院後 48 時間後または発症後 72 時間以内に抗菌薬を開始した RCT のみを集計)

95%信頼区間は、起こった結果 (死亡のオッズ比) の 95%が含まれる区間を示しています。例題の 1 つ目の論文 (Pederzoli et al) では、最も低いオッズ比は 0.12 で、最も高いオッズ比は 2.76 となり、どちらか一方で明らかに死亡が起こりやすいとは言えないこととなります。一方で、2 編目の Sainio et al の報告は、95%信頼区間が、オッズ比 1 をまたがずに全て「1」よりも左にあり、「統計学的に有意に左側 (抗菌薬側) のイベント (死亡) が少ない」と判定されます。

これを図で示したものが右側で、横棒の長さは 95%信頼区間を示し、その「横棒中央の小さな■」は個々の研究で報告されているアウトカム (ここでは致命率) の点推定値 (オッズ比, リスク比, 平均値など) を、示しています。なお、「横棒中央の小さな■」の大きさは、サンプルサイズ (症例数) を表し、症例数が多い場合には■が大きくなり、横棒は短くなり、信頼性が高くなります。

例題では、報告論文は 7 編です。95%信頼区間が、オッズ比 1 をまたがなければ、「統計学的有意差あり」と判定されますので、2 編目の Sainio et al の報告のみが、「有意差あり」、それ以外は有意差なしの判定となります。

(3) 統合結果

・メタ解析結果の読み方

7つの論文の「予防的抗菌薬の有無での致命率」のオッズ比を統合する方法が、図の上段に記載されています。「M-H, Random, 95%CI」と記載されているように、Mantel-Haenszel 検定を用いて、Random 効果モデルの設定で、95%信頼区間を提示する方式をとっています。

統合結果は、Total (95%CI) の行以下に記載されます。例題では、抗菌薬「Experimental」群の合計251例中22例が死亡「Events」、コントロール「Control」群では、244例中36例の死亡が示されています。「Weight」は、7編の各論文の結果に割り当てられた重み付けを示しています。統合計算は単純な相加平均ではなく、信頼区間の狭い報告（結果がばらついていない報告）程、より大きな重み付けが割り当てられて、統合されます。今回の例題の統合結果は、オッズ比0.60、その95%信頼区間は[0.33~1.07]と記載されています。これを右図で見ると、◆は統合結果の95%信頼区間を示しており、ひし形の左端は0.33、右端は1.07を指し示しています。この例題では、95%信頼区間の上限（ひし形の右端）は、1.07と「1」を超えているため、統合結果は統計学的に有意な結果を示していないと判定されます。

* 論文間の結論のばらつき（異質性）

次に、7編の論文間の結論のばらつき（異質性：Heterogeneity）について考えます。

図1の下から2行目、Heterogeneityにおける記載を見ると、Tau²乗検定で0.00、カイ²乗検定で2.87、 $P=0.83$ と計算され、有意($P<0.05$)な異質性は無しと判定されます。I²乗検定でも異質性は0%と判定されています。つまり、7編のRCTは、予防的抗菌薬の有無での致命率に関して結論が同じであるといえます。

なお、7つの論文の中で、2つ目の論文(Sainio et al)だけが有意な結論を示しているため、7つの論文間には異質性があると考えられるかもしれません。しかし、2つ目の論文は抗菌薬群30例中死亡1例、コントロール群30例中死亡7例のRCTで、95%信頼区間が[0.01~0.99]と広く、メタ解析で割り当てられた重み付け「Weight」は7.2%に過ぎず、報告結果の重要性が乏しいと考えられ、全体の異質性にはあまり影響していない評価となります。

* メタ解析による総合評価

最下段の本検討の総合評価（Test for overall effect）は、 $P=0.08$ と記載されていることから、今回の仮説「予防的抗菌薬を使うと、抗菌薬なしに比べて致命率は改善する」には、有意差($P<0.05$)は認められない、という判定となりました。

2) エビデンス総体

以前は各研究の表1の研究のデザインでエビデンスの確かさ（確実性）を評価していました。しかし、近年は、「致命率」「感染性合併症」などのアウトカムごとに、エビデンスの確かさを上記のようなForest plotを作成し評価します。その際に、表3の「エビデンスレベルを下げる要因」や「エビデンスレベルを上げる要因」がある場合には、表4のエビデンスの確か性を1ランクまたは2ランク、下げたり、上げたりします。各アウトカムの臨床上的重要性を重み付けし、全てのアウトカムでの知見（「エビデンス総体」）から、エビデンスの確か性（表4）が定まります。



※表1, 3, 4は次頁に掲載

表 1 研究デザイン分類

各文献へは下記7種類の「研究デザイン」を付記した。	
CPG	Clinical practice guidelines：診療ガイドライン
SR	Systematic review：システマティックレビュー
MA	Meta-analysis：RCTのメタ解析
RCT	Randomized controlled trial：ランダム化比較試験
OS	Observational study, Cohort study, Case control study, Cross sectional study： 観察研究, コホート研究, 症例対照研究, 横断研究
CS	Case series, Case study：症例集積研究, 症例報告
EO	Expert opinion：専門家の意見*
(*患者データに基づかない専門委員会や専門家個人の意見は、本ガイドラインではEOとして作成の参考にしたが、エビデンスとしては用いないこととした)	

(文献 14 より引用改変)

表 3 アウトカムごと、研究デザインごとの蓄積された複数論文の総合評価

<p>1) 初期評価：各研究デザイン群の評価</p> <p>SR (システマティックレビュー), MA (メタ解析), RCT 群 = 初期評価「高」</p> <p>OS (観察研究) 群 = 初期評価「低」</p> <p>CS (症例集積, 症例報告) 群 = 初期評価「非常に低」</p> <p>2) エビデンスレベルを下げる要因の有無の評価</p> <p>研究の質にバイアスリスクがある (表 2 の結果)</p> <p>結果に非一貫性がある……………複数の論文間で結論が異なる</p> <p>エビデンスの非直接性がある…論文内容と CQ 間でずれがある, または論文内容を, 日本の臨床にそのまま適応できない (医療保険等)</p> <p>データが不精確である……………症例数が不十分, または予定例数に到達しない</p> <p>出版バイアスの可能性が高い…都合のいい結果のみが報告されている</p> <p>3) エビデンスレベルを上げる要因の有無の評価</p> <p>大きな効果があり, 交絡因子がない…全例に大きな効果が期待される</p> <p>用量-反応勾配がある……………用量を増やせば, さらなる効果が期待できる可能性のある交絡因子が, 真の効果をより弱めている</p> <p>総合評価：最終的なエビデンスの確実性「高, 中, 低, 非常に低」を評価判定した。</p>
--

(文献 25 より引用改変)

表 4 エビデンスの確実性

<p>高：質の高いエビデンス</p> <p>真の効果がその効果推定値に近似していると確信できる。</p> <p>中：中程度の質のエビデンス</p> <p>効果の推定値が中程度信頼できる。</p> <p>真の効果は, 効果推定値におおよそ近いが, 実質的に異なる可能性もある。</p> <p>低：質の低いエビデンス</p> <p>効果推定値に対する信頼は限定的である。</p> <p>真の効果は, 効果の推定値と, 実質的に異なる可能性がある。</p> <p>非常に低：非常に質の低いエビデンス</p> <p>効果推定値がほとんど信頼できない。</p> <p>真の効果は, 効果の推定値と実質的にのおおよそ異なる可能性が高い。</p>
--

(文献 25 より引用改変)

※表 2 および引用文献 14, 25 は本書第 1 章に掲載