

メタアナリシスにおける平均治療効果の推定

野間 久史 データ科学研究系 助教

【メタアナリシスにおける試験間異質性と平均治療効果】

メタアナリシスとは、過去に行われた臨床試験の結果を統合し、関心のある薬剤・治療法の治療効果の大きさを総合的に評価するための研究手法である。一般的に、統合を行う試験の間では、セッティング・対象者の背景・介入方法・アウトカムの定義などが異なり、対象となる試験間で「治療効果の大きさがまったく同じ」と見なすことは難しい。したがって、以下の変量効果モデル (DerSimonian and Laird, 1986) を用いて、平均的な治療効果の評価するのが、現在の標準的な方法となっている。

$$Y_k \sim N(\theta_k, \sigma_k^2) \quad (k=1, 2, \dots, K)$$

$$\theta_k \sim N(\mu, \tau^2)$$

θ_k : True treatment effect of k -th trial (e.g., log odds-ratio)
 Y_k : Estimator of θ_k
 σ_k^2 : Variance of Y_k
 μ : Average treatment effect
 τ^2 : Heterogeneity variance

しかし、変量効果の分布の平均 μ を関心のある平均治療効果の指標とすることには、以下のような問題が考えられる。

1. (統計的な問題) μ の推定には、統合する試験の数 (K) が無限大というもとの大標本近似を用いざるを得ない。一般的に、医学研究のメタアナリシスでは、 $K < 20$ であり、この近似は成り立たない。実際、近年の研究により、この大標本近似に基づく既存の信頼区間は、実践的な設定のもとで、真値の被覆確率が名目水準を下回ることが知られている (Brockwell and Gordon, 2001; Noma, 2011)。つまり、統計的な誤差を過小評価してしまう。 K が5以下のもとでは、特に深刻であるが、現実的には、後に示す Chan *et al.* (2007) の事例のように、一般にも多くあり得る設定である。
2. (臨床的な解釈の問題) メタアナリシスの結果で関心があるのは、「試験間の異質性があるもとの平均的な治療効果を要約した指標」である。上記の変量効果モデルの「正規分布の平均 μ 」は、この目的に適ったものであるのか？ そもそも、この指標は、上記の数学的な仮定が完全に正しいもとの「平均治療効果」と解釈することができる。変量効果の分布の正規性の仮定は、数学的な利便性のための仮定であり、多くの医学研究者にとっては、臨床的な根拠や意義はない。つまり、この正規性の仮定が成り立たない場合、 μ のものが平均治療効果としての解釈ができず、その推定量がなにを推定しているのかさえ、わからない。

“Do the random-effects models give a fairly precise answer?”

【新しい平均治療効果の指標の提案と統計的推測】

上記のような問題を解決するためのひとつのアプローチとして、本研究では、新しい平均治療効果の指標を提案する。モデルには、変量効果モデルを利用せず、

$$Y_k \sim N(\theta_k, \sigma_k^2) \quad (k=1, 2, \dots, K)$$

θ_k s are constants, and possibly not equal.

というモデルを仮定する。平均治療効果は、個々の試験ごとの「真の治療効果 θ_k 」を、適当な試験ごとの重みによって重みつけた「重みつき平均」として定義する。すなわち、

$$\delta = \frac{\sum_{k=1}^K w_k \theta_k}{\sum_{k=1}^K w_k}$$

として、「試験ごとの真の治療効果を平均したもの」として、定義する。重み w_k は、研究者や意思決定者が任意に定めることができるが、以下のようなものが代表的なものとして考えられるだろう。

- (1) δ_I : Equal weights; $w_k = 1$
- (2) δ_{II} : Proportional to sample sizes; $w_k = N_k$ (Total sample size of k th trial)
- (3) δ_{III} : Proportional to precision (inverse variance) $w_k = 1/\sigma_k^2$

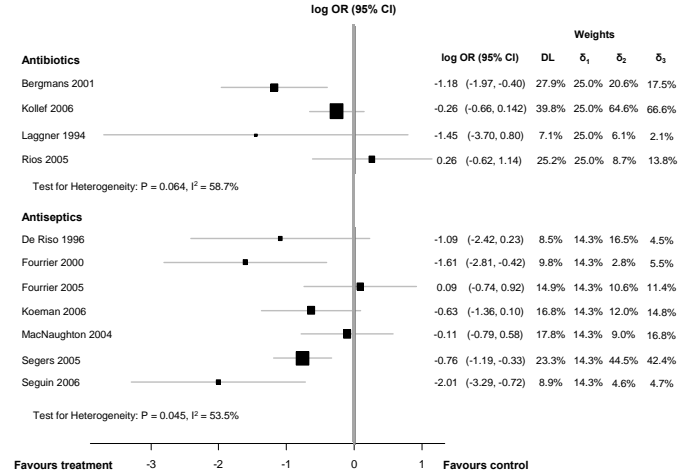
平均治療効果の指標 δ は、上記の Y_k のモデルにのみ依存しており、変量効果モデルなどの便宜的な強い数学的仮定に依存しない。 δ の推測においては、

$$\hat{\delta} = \frac{\sum_{k=1}^K w_k Y_k}{\sum_{k=1}^K w_k}, \text{Var}(\hat{\delta}) = \frac{\sum_{k=1}^K w_k^2 \sigma_k^2}{\left(\sum_{k=1}^K w_k\right)^2}$$

を用いればよい。古典的な推定方程式の議論から、一致性、漸近正規性が成り立つことは示すことができる。検定・信頼区間の構成においては、Wald型の近似手法を用いればよい。ただし、このもとの必要となる大標本の仮定は「 $K \rightarrow \infty$ 」ではなく、「 $N_k \rightarrow \infty$ ($k=1, 2, \dots, K$)」である。これは、メタアナリシスの設定でも理に合ったものと言える。

【参考文献】

[1] DerSimonian R, Laird NM. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; 7: 177-188.
 [2] Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. *Statistics in Medicine* 2001; 20: 825-840.



Chan *et al.* (2007) による人工呼吸器除染による肺炎予防についてのメタアナリシス。

【人工呼吸器関連肺炎の予防についてのメタアナリシス】

Chan *et al.* (2007) は、成人を対象とした人工呼吸器による肺炎の予防についてのメタアナリシスを行っている。対象となったのは、抗生物質 (Antibiotics) による口腔内除菌に関するランダム化比較試験 4試験1098人、消毒薬 (Antiseptics) による口腔内除菌のランダム化比較試験 7試験2144人である。治療効果の異質性の検定は、それぞれ $P = 0.064, 0.045$ であり ($I^2 = 58.7\%, 53.5\%$)、治療効果の異質性が示唆されている。原著での解析には、変量効果モデルが用いられており、肺炎の発生に関するオッズ比の推定値は、それぞれ 0.69 (95% CI: 0.41, 1.18), 0.56 (95% CI: 0.39, 0.81) となっている。試験数 K が小さく、特に抗生物質の試験については4試験しかなく、変量効果モデルによる推測は本質的に不適切であるともいえる。一方で、抗生物質の効果については有意な結果が得られておらず、原著の結論でも「More evidence is needed before firm conclusions can be made on the effect of antibiotic oral decontamination.」と述べられている。

以下は、古典的な固定効果モデル (Mantel-Haenszel法)、変量効果モデル (DerSimonian-Laird法、最尤法)、提案する平均治療効果の推定の結果をまとめたものとなる。試験間の異質性が強く示唆されているため、Mantel-Haenszel法による共通オッズ比の推定結果の妥当性はあまり見込めない。また、変量効果モデルによる推定結果には、上記のような問題もあり、妥当性が疑われる。一方、提案する平均治療効果の推定では、いずれも信頼区間は変量効果モデルよりも狭くなっており、抗生物質についても有意な結果が得られている。導かれる結論が逆転することになるが、提案法の信頼区間の妥当性と精度の改善は、シミュレーションの結果でも確認されている。これらの結果は、臨床家にとっても、興味深いものとなると思われる。

Odds Ratio (95% CI)				
Model	Target parameter	Estimator	Antibiotics ($K = 4$)	Antiseptics ($K = 7$)
Fixed-effects model	Common OR	Mantel-Haenszel	0.667 (0.484, 0.919)	0.503 (0.383, 0.661)
Random-effects model	μ : grand mean	DerSimonian-Laird	0.625 (0.326, 1.197)	0.488 (0.308, 0.772)
Random-effects model	μ : grand mean	Maximum Likelihood	0.641 (0.267, 1.296)	0.500 (0.273, 0.782)
Proposed model	δ_I	$\hat{\delta}_I$	0.518 (0.272, 0.985)	0.417 (0.288, 0.604)
Proposed model	δ_{II}	$\hat{\delta}_{II}$	0.621 (0.440, 0.875)	0.481 (0.346, 0.668)
Proposed model	δ_{III}	$\hat{\delta}_{III}$	0.688 (0.495, 0.954)	0.519 (0.392, 0.687)

[3] Noma H. Confidence intervals for a random-effects meta-analysis based on Bartlett-type corrections. *Statistics in Medicine* 2011; 30: 3304-3312.
 [4] Chan EY, Ruest A, Meade MO. Oral decontamination for prevention of pneumonia in mechanically ventilated adults: systematic review and meta-analysis. *British Medical Journal* 2007; 334: 889.