メタ・アナリシス—エビデンスの統合をめざす

丹 後 俊 郎*

Meta-Analysis: Aiming to Integrate the Evidence

Toshiro TANGO

Meta-analysis is a statistical method which examines all the combinable studies on a particular subject conducted independently in the past in order to estimate the size of the common effect regarding the same research question, e.g., to estimate the efficacy, toxicity effect and environmental risk of certain treatments. In other words, meta-analysis is a scientific procedure that integrates the evidence obtained all over the world. This paper describes its application to clinical research.

Key words: effect size, cumulative meta-analysis, publication bias, fixed-effects model, random-effects model, Bayesian

1 はじめに

過去に行われた研究結果をまとめようとする努力はそんなに新しいことではない。1896年、イギリスのSir Wrightは腸チフスに対するワクチンを開発し、その効果を幾つかのグループで検討した。この結果、ある程度の効果が認められたとして、このワクチンをイギリス軍の兵士の予防のためにルーチンで使用することを薦めたのである。その同じ年、統計学者Karl Pearson (1904)はこのワクチンの採用に関してそれまでに試された効果のデータを再検討することを依頼されたのである。腸チフスによる死亡と予防接種の関連を検討した6つの調査結果からワクチンの効果を評価するために、それぞれの調査結果に四分相関係数 (tetra-choric correlation) を計算し、6つの相関係数の平均

値 ($\pm$ S.E.) を求めたのである。その主要な目的は「それぞれのサンプルが小さくて、それぞれの結果からでは明確なエビデンス (evidence) が得られない」からである。これは現在のメタ・アナリシスを行う主要な動機である。この結果から、彼は「相関係数があまりにも低いのでワクチンを採用することは薦められない、むしろ、相関係数がより高くなるようにワクチンの改良研究をすべきだ」と結論したのである。メタ・アナリシスを実施した最初の研究者がこのPearsonであった。時代がすこし進んで、農業研究の盛んな時代になると、古典的な統計学のテキスト、たとえばTippet (1931) の *The Methods of Statistics*, または、Fisher (1932) の *Statistical Methods for Research Workers* の中に、同じテーマに関して独立に行われた複数の研究結果を統合する統計的方法の解説が初めて登場してくる。

一方、治療効果における最初の統合の試みはおそらくBeecher (1955) による偽薬、プラセボ (placebo) の効果であろう。その当時は、術後の傷の痛み、咳嗽、狭心症による痛み、頭痛など、様々な症状を緩和する目的として投与されていたプラセボ (Agentと静脈注射、皮下注射、経口投与と投与方法はちがうものの) の効果を評価したものである。彼は15の試験を無作為に選んで (実際にはもっと選択できたと述べている) 有

国立公衆衛生院疫学部 (2002年3月まで)
国立保健医療科学院 技術評価部 (2002年4月より)
〒351-0197 埼玉県和光市南2丁目3-6
Professor of Biostatistics
Director, Department of Biostatistics & Technology
Assessment, National Institute of Public Health, 3-6
Minami 2 Chome, Wako, Saitama-ken 351-0197
Japan.
Tel: (048)458-6111, ext. 2711
Fax: (048)469-3875
Email: tango@iph.go.jp

効率の平均 ($\pm$ SE) を計算してみると $35.2 \pm 2.2\%$ の患者に症状の緩和がみられたという結果であった。論文のタイトルも The Powerful Placebo であった。

もっとも、メタ・アナリシスという言葉の誕生とともに、メタ・アナリシスという言葉の誕生とともに、社会科学の世界、特に、教育学の研究においてである。その背景には、その当時、いろいろな教育法に関する数多くの研究が存在し、その要約を目的とした記述的なレビューでは文献の選択の方法とそれぞれの研究結果の評価の仕方が reviewer の興味、専門性、などに強く依存し、極めて主観的になってしまうという危機感があったのである。心理学者 Glass (1976) が始めて meta-analysis という名称を提案した。この分野の集大成としてまとめられたものが Cooper and Hedges (1994) による The Handbook of Research Synthesis である。

臨床試験の世界に meta-analysis が持ち込まれたのはイギリスの巨人 Richard Peto の存在が大きい。Peto のデザインによる Yusuf et al. (1985) の心筋梗塞後の β ブロッカーの長期投与の 2 次予防効果のメタ・アナリシスはあまりにも有名である。これを契機に臨床試験の評価にメタ・アナリシスが急速に広がっていく。早期乳癌治療合同研究班 (Early Breast Cancer Trialists' Collaborative Group, 1988) による早期乳癌のアジュバント化学療法、Iain Chalmers et al. (1989) による周産期領域における治療効果に関するメタ・アナリシスは代表的なものである。

2. メタ・アナリシスの基本的手順

きちんとデザインされ、計画通りに実施されたメタ・アナリシスは個々の研究結果を記述的にレビューしたり単純に平均された推定値よりはるかに客観的で正確な効果・リスクの推定値が得られる。しかし、その実施にあたっては、「統合可能な研究群」をどう選ぶかが大問題であり、他の研究と同様に慎重な研究プロトコルを事前に作成する必要がある。そのポイントは次の手順に整理できる。

1. 選択基準 (研究の質、研究デザイン、研究規模など) の明確化
2. 文献の網羅的探索 (publication bias をさけるための方法を工夫する)
3. 効果・リスク指標の選択
 - (a) 計量値であれば「差とその標準偏差」

- (b) 2 値 (binary) であればオッズ比がよく利用される (効果の差 (%)) はあまり利用されない。

4. 統合可能性 (均質性) の検討

5. 統計手法の選択 (統合可能性の検定も含めて)

- (a) 母数モデル (fixed-effects model): 研究間の結果の差はもっぱら偶然変動だけであるという均質性 (homogeneity) を仮定する方法で、Mantel-Haenszel の方法、Peto の方法などが代表的である。
- (b) 変量モデル (random-effects model): 研究間には偶然変動としての誤差以外にも無視できない異質性 (heterogeneity) があると考えこれを確率変数としてモデル化した方法で、DerSimonian-Laird の方法などが代表的である。したがって、母数モデルより若干信頼区間の幅が広がる。
- (c) ベイズモデル (Bayesian model): 効果・リスクの大きさは一定の値ではなく、ある確率分布 (事前分布) をしていると「信じる」方法で、その「信念」をデータから計算される「事後分布」で更新させる方法である。最近の流行でもある。

6. 感度分析 (sensitivity analysis) の実施

3. 心筋梗塞後の 2 次予防への β blockers

図 1 には心筋梗塞患者の 2 次予防のための β ブロッカー長期投与の効果 (死亡率リスクの減少) を検討した 17 の臨床試験のメタ・アナリシスの例を示す。試験毎に薬剤 β ブロッカーの死亡オッズ比の点推定値 ($\square$ で示す) と 95% 信頼区間が示されている (薬剤の効果があれば死亡オッズ比が小さくなる)。信頼区間が 1.0 を含んでいればその試験での治療効果は有意でなかったことを示している。また、四角形の面積はメタ・アナリシスでの重みを意味し、それはだいたい、標本サイズに比例 (信頼区間の幅に反比例) する。つまり、規模の小さい研究結果の重みは小さい。さて、2 つの試験を除いて他の 15 の試験では治療効果は有意ではない。しかし、Peto の方法で統合されたオッズ比は ($\blacksquare$ で示す) 0.78 (95% CI: 0.71–0.87) であり、死亡オッズは死亡相対リスクに近似できるので 22% ($1 - 0.78 = 0.22$) のリスク減少が期待できる、という解釈が可能である。17 すべての試験での信頼区間がこの

Peto's method

図1. 心筋梗塞患者の2次予防のための β ブロッカー長期投与の効果(死亡率リスクの減少)を検討した17の臨床試験のメタ・アナリシス。

併合オッズ比 0.78 (図1の点線)を含んでいるので、17の試験はかなり同質であることを示唆している。事実、統合可能性を検定してみるとその p 値=0.2で否定できない。

しかし、これで解析が終わったわけではない。メタ・アナリシスはいくらプロトコルで条件を設定したといっても、過去に行われた研究を検討するのであるから、いろいろな意味での不安定要因が多く、条件設定の妥当性を検討することは不可能に近い。したがって、条件を変えて計算を繰り返したときに結果がどの程度変わるかを検討しておくことは極めて重要である。これを感度分析という。統計手法、試験の質(例: randomization, エンドポイントの評価方法、解析方法などで総合的に判断)、試験の規模などで分類して計算を繰り返すことが重要である。特に試験の規模で分類することにより「有意な結果だけが公表された傾向があった否か」という publication bias をチェックできる可能性がある。つまり、同じ効果が推定されても規模の小さい試験では有意になりにくいのであるから、

もし publication bias があれば、試験規模が小さいほど推定された効果が大きい、言い換えれば、最も大きな試験が最も小さい効果を示しているはずである。図2に β ブロッカーの感度分析の例(Egger et al, 1997)を引用したが、試験規模での層別解析ではやはり100人以上の死亡が観察された試験群で推定された効果が最も小さく、publication bias の存在が疑われる。しかし、全体的には大きな影響はないようである。

4 Joseph Lau の累積メタ・アナリシス

メタ・アナリシスを適用し解釈する上で画期的な方法が Joseph Lau et al. (1992) により提案された。それは、なにも難しいことではなく、「新しい RCT の報告がでるたびにメタ・アナリシスを繰り返してその結果を図示していく」簡単な方法であり、累積メタ・アナリシス(cumulative meta-analysis)と命名された。これを利用すれば「初めて有意な効果に達した年代を遡って同定できる」利点がある。Lau et al. (1992) は

図 2. β ブロッカーのメタアナリシスの感度分析の例 (Egger et al., 1997).

急性心筋梗塞後の血栓溶解剤としての streptokinase の静脈内投与の有効性に関する 33 件の RCT のメタ・アナリシスにおいて、Mantel-Haenszel 法を用いた累積メタ・アナリシスを適用した(図 3 参照)。その結果、

1. 1971 年の段階で初めて治療効果が有意となった ($p=0.023$)
2. 1973 年には、それまで 8 つの試験で総勢 2,432 人の患者が無作為割付されていて、治療効果の死亡オッズ比は 0.74 (95% CI: 0.59-0.92), $p=0.0071$ となり、この有効性の評価はこの時点から不変である。
3. 1977 年には p 値が更に小さくなり $p<0.001$,
4. これらのメタ・アナリシスの結果は 1986 年, 1988 年に Lancet に報告された大規模試験 GISSI (the Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico) と ISIS-2 (the Second International Study of Infarct

Survival) の結果によっても変わらなかった。それは、単に p 値, 信頼区間の幅を更に小さくしてその有効性の印象を強めるのに役立っただけである。

5. 1988 年には死亡オッズ比の推定値が 0.77 ($p<10^{-15}$) となった。つまり、23% の死亡オッズ(リスク) 減少が期待されるというエビデンスが得られる。

ということを示した。ところで、この治療法が認可された時期は国によって異なるのが面白い。ドイツだけは二つの大規模試験の報告の前 1985 年に認可しているが、GISSI の報告を待って認可した国がほとんどで、1986 年にはイタリア、ニュージーランド、オランダ、スウェーデン、1987 年にはメキシコ、米国、スイスで認可されている。しかしノルウェー (1988)、オーストラリア (1988)、フランス (1988)、イギリス (1989) は SISI-2 の結果を待って認可したといった具合である。つまり、認可される 10 年以上も前にその有効性は立証されていたことを累積メタアナリシスは示しているのである。

このように、メタ・アナリシスはエビデンスに基づく治療法の推薦、最新の治療ガイドラインなどにその利用が期待される。

5. 基本的なモデル

ここではメタ・アナリシスの基本モデルである母数モデルについて解説する。具体的な計算プロセスを示す目的と臨床試験へのメタ・アナリシスの伝道者である Peto に敬意を表して彼の方法をまず紹介し、その次に一般モデルを解説する。

5.1 臨床試験のための Peto の方法

Peto の方法 (Yusuf et al., 1985) は臨床試験 RCT の effect size としてオッズ比を用いてメタ・アナリシスを行うために提案された方法であり、それぞれの RCT (Randomized Controlled Trial) の結果の解釈を容易にする統計量で構成されている。データは表 1 に示す 2×2 分割表である。この方法は各試験毎に、治療群における観測イベント数から治療効果がない場合に期待されるイベント数を引いた数、つまり、その試験でイベントの発生を予防できた患者の数を表す

$$O_i - E_i, \quad i=1, 2, \dots, K$$

図3. 急性心筋梗塞後の血栓溶解剤としてのstreptokinaseの静脈内投与の有効性に関する33件のRCTのMantel-Haenszel法を用いた累積メタ・アナリシス (Lau et al., 1992)

表1. 臨床試験における原因と結果の K 個の研究結果の 2×2 分割表, $i=1, \dots, K$.

原因	結果	事象の発生		計
		あり	なし	
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

とその分散 V_i から近似的に推定されるオッズ比

$$\hat{OR}_i = \frac{a_i d_i}{b_i c_i} \cdot \exp\left(\frac{O_i - E_i}{V_i}\right)$$

を利用して計算するものである。その計算のプロセスは表2に示す。

ただ、Petoの方法はあくまで無作為割付に基づく臨床試験のデータに適用するものである。疫学研究の

データには適用すべきではない (unbalanced data によるバイアスが生じる)。

5.2 漸近的正規近似に基づく一般的なモデル

一般に、実験 (曝露) 群 (E: experimental group, exposed group) の治療成績 (曝露リスク) を対照 (非曝露) 群 (C: control group, unexposed group) の治療成績 (曝露リスク) と比較する指標 (有効率の差, オッズ比など) の大きさ effect size を θ とし, メタ・アナリシスの対象となる研究の数が K 個あるとする。各研究からコピーあるいは計算した effect size の推定値を

$$(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_K) \quad (1)$$

としよう。各研究でのサンプルサイズが大きければ, $\hat{\theta}$ には漸近的に最良な最尤推定量を考えることがで

表2 母数モデル：Peto の方法

-
1. 各研究での対数オッズ比の近似値を計算する

$$\log \bar{O}\hat{R}_i = (O_i - E_i) / V_i = \left(a_i - \frac{n_{1i}m_{1i}}{n_i} \right) / \left(\frac{n_{1i}m_{0i}m_{10}m_{0i}}{n_i^2(n_i - 1)} \right)$$

2. 対数オッズ比の標準誤差を計算する

$$SE_i = 1 / \sqrt{V_i}$$

各研究の 95% 信頼区間は $\exp(\log \bar{O}\hat{R}_i \pm 1.96 SE_i)$ で計算する。

3. 各研究の重みを計算する

$$w_i = 1 / SE_i^2 = V_i$$

4. 統合オッズ比の推定：重み付き平均を計算し指数変換でもとに戻す。

$$\bar{O}\hat{R}_p = \exp \left(\frac{\sum_{i=1}^K (O_i - E_i)}{\sum_{i=1}^K V_i} \right)$$

5. 統合オッズ比の 95% 信頼区間を計算する

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

6. 均質性の検定を行う。

$$Q_1 = \sum_{i=1}^K \frac{(O_i - E_i)^2}{V_i} - \frac{\sum_{i=1}^K (O_i - E_i)^2}{\sum_{i=1}^K V_i} \sim \chi_{K-1}^2$$

7. 有意性の検定を行う (Mantel-Haenszel 検定)

$$Q_2 = \frac{\{|\sum_{i=1}^K (O_i - E_i)| - 0.5\}^2}{\sum_{i=1}^K V_i} \sim \chi_1^2$$

きる (例えば, 2×2 分割表でのそれぞれのセルで 5 例以上). そこで, 適当な変換 $f(\theta)$ により漸近的正規近似

$$f(\hat{\theta}_i) | \theta_i, s_i^2 \stackrel{\text{asympt}}{\sim} N(f(\theta_i), s_i^2), \quad i=1, 2, \dots, K \quad (2)$$

が仮定できる状況を考える. ここで, s_i^2 は $f(\hat{\theta}_i)$ の漸近分散で既知とするが, 後の議論のために

$$w_i = \frac{1}{s_i^2} \quad i=1, \dots, K \quad (3)$$

と重み (weight) を表す変数 w_i を導入しておく.

さて, 式 (2) の条件において, 一つの自然なメタ・アナリシスでの帰無仮説は,

$$H_0: \theta_1 = \dots = \theta_K = \theta \quad (4)$$

であろう. このモデルは, 各研究結果は同一の effect size θ を持つ点で均質性を仮定した方法といえる. 言い換えれば, 未知の effect size θ を推定すべき母数として各研究結果がその回りに分布している確率変数であるという母数モデルである.

帰無仮説 H_0 の下では $f(\theta)$ の対数尤度 $l(f(\theta)) =$

$l(f(\theta) | f(\hat{\theta}_i), s_i^2)$ は

$$l(f(\theta)) \propto Q = \sum_{i=1}^K \left(\frac{f(\hat{\theta}_i) - f(\theta)}{s_i^2} \right)^2 = \sum_{i=1}^K w_i (f(\hat{\theta}_i) - f(\theta))^2 \quad (5)$$

となるので, $f(\theta)$ の漸近的最尤推定量は

$$f(\hat{\theta})_{\text{AMLE}} = \frac{\sum_{i=1}^K f(\hat{\theta}_i) w_i}{\sum_{i=1}^K w_i} \quad (6)$$

となる. つまり,

$$\hat{\theta}_{\text{AMLE}} = f^{-1} \left(\frac{\sum_{i=1}^K f(\hat{\theta}_i) w_i}{\sum_{i=1}^K w_i} \right), \quad (7)$$

$$95\% \text{ CI}: f^{-1} \left(f(\hat{\theta})_{\text{AMLE}} \pm 1.96 \sqrt{\frac{1}{\sum_{i=1}^K w_i}} \right) \quad (8)$$

となる. なぜなら

$$\text{Var}(f(\hat{\theta})_{\text{AMLE}}) = \frac{1}{\sum_{i=1}^K w_i} \quad (9)$$

が成立する. ところで,

$$\begin{aligned}
Q &= \sum_{i=1}^K \left(\frac{f(\hat{\theta}_i) - f(\theta)}{s_i^2} \right)^2 \\
&= \sum_{i=1}^K w_i (f(\hat{\theta}_i) - f(\theta))^2 \stackrel{\text{asympt}}{\sim} \chi_K^2 \\
&= \underbrace{\sum_{i=1}^K w_i (f(\hat{\theta}_i) - f(\hat{\theta})_{\text{AMLE}})^2}_{\chi_{K-1}^2} \\
&\quad + \underbrace{\sum_{i=1}^K w_i (f(\hat{\theta})_{\text{AMLE}} - f(\theta))^2}_{\chi_1^2}
\end{aligned} \quad (10)$$

と分解できる。前者の統計量は帰無仮説 H_0 の検定, つまり, 各研究での治療効果の均質性の検定統計量である。メタ・アナリシスでは研究の統合可能性 (combinability) の検定とも言う。後者は均質性の仮定の下で $H_0: \theta = \theta_0$ の検定統計量になる。特に, 「効果がない」とする帰無仮説 $H_0: f(\theta) = 0$ の有意性検定が興味ある。

$$\text{均質性の検定} \Rightarrow Q_1 = \sum_{i=1}^K w_i (f(\hat{\theta}_i) - f(\hat{\theta})_{\text{AMLE}})^2 \sim \chi_{K-1}^2 \quad (11)$$

$$\text{有意性の検定} \Rightarrow Q_2 = \frac{(\sum_{i=1}^K w_i f(\hat{\theta}_i))^2}{\sum_{i=1}^K w_i} \sim \chi_1^2 \quad (12)$$

6. Heterogeneity をモデルする

6.1 変量モデル

母数モデルでは θ_i は共通 (均質) と考えたが, これは現実を少々単純化しすぎたモデルであり, 現実には各研究結果 $\hat{\theta}_i$ は本質的にはある程度の差がある (プロトコルの違い, 患者の違い, 地域の違い, 研究者の違いなど) と考える方が自然であろう。そこで, この違い, すなわち異質性 (heterogeneity) をモデル化した一つの自然なモデルとして

$$f(\theta_i) | \theta, \tau^2 \sim N(f(\theta), \tau^2), \quad i=1, 2, \dots, K \quad (13)$$

という変量モデル (random-effects model) を考えることができる。この仮定の下では, 式 (2) は

$$f(\hat{\theta}_i) | \theta_i, s_i^2, \tau^2 \stackrel{\text{asympt}}{\sim} N(f(\theta_i), s_i^2 + \tau^2), \quad i=1, 2, \dots, K \quad (14)$$

と置き換えられる。変量モデルでは $f(\theta), \tau^2$ の周辺尤度を最大化する制限付最尤推定量 (REML, restricted maximum likelihood estimator) を考えるのが自然である。その対数尤度 $l(f(\theta), \tau^2) = l(f(\theta), \tau^2 | f(\hat{\theta}_i), s_i^2)$ は

$$\begin{aligned}
l(f(\theta), \tau^2) &\propto \sum_{i=1}^K \left(\frac{(f(\hat{\theta}_i) - f(\theta))^2}{s_i^2 + \tau^2} + \log(s_i^2 + \tau^2) \right) \\
&\quad + \log \left(\sum_{i=1}^K \frac{1}{s_i^2 + \tau^2} \right)
\end{aligned} \quad (15)$$

となるので, 重み変数を

$$w_i(\tau) = \frac{1}{s_i^2 + \tau^2} \quad i=1, \dots, K \quad (16)$$

とおくと,

$$\hat{\theta}_{\text{REML}} = f^{-1} \left(\frac{\sum_{i=1}^K f(\hat{\theta}_i) w_i(\hat{\tau})}{\sum_{i=1}^K w_i(\hat{\tau})} \right), \quad (17)$$

$$95\% \text{ CI: } f^{-1} \left(f(\hat{\theta})_{\text{REML}} \pm 1.96 \sqrt{\frac{1}{\sum_{i=1}^K w_i(\hat{\tau})}} \right) \quad (18)$$

となる。ここで, $\hat{\tau}^2$ は次の非線形方程式の解となる:

$$\hat{\tau}^2 = \frac{\sum_{i=1}^K w_i^2(\hat{\tau}) \left(\frac{K}{K-1} (f(\hat{\theta}_i) - f(\hat{\theta})_{\text{REML}})^2 - s_i^2 \right)}{\sum_{i=1}^K w_i^2(\hat{\tau})} \quad (19)$$

これは繰り返し収束計算が必要である。一方, 均質性の検定統計量 Q_1 を利用したモーメント法を適用するとより繰り返し計算の必要がない推定値が得られる。先ほどの分解の逆を考えると

$$\begin{aligned}
Q_1 &= \sum_{i=1}^K w_i (f(\hat{\theta}_i) - f(\hat{\theta})_{\text{AMLE}})^2 \\
&= \sum_{i=1}^K w_i (f(\hat{\theta}_i) - f(\theta))^2 - \left(\sum_{i=1}^K w_i \right) (f(\hat{\theta})_{\text{AMLE}} - f(\theta))^2
\end{aligned}$$

となるから

$$\begin{aligned}
E(Q_1) &= \sum_{i=1}^K w_i \text{Var}(f(\hat{\theta}_i)) - \left(\sum_{i=1}^K w_i \right) \text{Var}(f(\hat{\theta})_{\text{AMLE}}) \\
&= \sum_{i=1}^K w_i \left(\frac{1}{w_i} + \tau^2 \right) \\
&\quad - \left(\frac{1}{\sum_{i=1}^K w_i} + \frac{\tau^2 \sum_{i=1}^K w_i^2}{\left(\sum_{i=1}^K w_i \right)^2} \right) \left(\sum_{i=1}^K w_i \right) \\
&= (K-1) + \tau^2 \left(\sum_{i=1}^K w_i - \frac{\sum_{i=1}^K w_i^2}{\sum_{i=1}^K w_i} \right)
\end{aligned} \quad (20)$$

となる。つまり, モーメント推定量が次式で計算できる。

$$\hat{\tau}^2 = \max \left\{ 0, \frac{Q_1 - (K-1)}{\sum_{i=1}^K w_i - \frac{\sum_{i=1}^K w_i^2}{\sum_{i=1}^K w_i}} \right\} \quad (21)$$

これが DerSimonian and Laird (1986) の方法である。もし, $\hat{\tau}^2 = 0$ の場合は母数モデルに一致する。これは $Q_1 < K-1$ の時に起こるが, それは $\tau^2 = 0$ のとき

$E(Q_1)=K-1$ であるから母数モデルが選ばれて妥当な結果である。

6.2 Bayesian モデル

式 (13) の変数モデルをもう一度考えてみよう。

$$f(\theta_i) | \theta, \tau^2 \sim N(f(\theta), \tau^2), \quad i=1, 2, \dots, K$$

このモデルでは、パラメータ θ, τ^2 (超パラメータ, hyperparameters という) は定数と考えており、その不確実性を考慮していない。それを考慮するためには full Bayes 法を適用すれば良い。たとえば、超パラメータの事前分布として無情報事前分布 (noninformative prior)

$$f(\theta) \sim N(0, a) \quad a=100 \text{ (程度)} \quad (22)$$

$$1/\tau^2 \sim \text{Gamma}(a, a) \quad a=0.001 \text{ (程度)} \quad (23)$$

とする。ここに、Gamma(a, b) はガンマ分布でその確率密度関数は

$$g(x | a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx) \propto x^{a-1} \exp(-bx) \quad (24)$$

$$E(X) = \frac{a}{b}, \quad \text{Var}(X) = \frac{a}{b^2} \quad (25)$$

である。そうすると、 $\{f(\theta), f(\theta_1), \dots, f(\theta_K), \tau^2\}$ に関する事後分布は

$$p(f(\theta), f(\theta_i), \tau^2 | \hat{\theta}_i, s_i^2) \\ \propto \prod_{i=1}^n p(f(\hat{\theta}_i) | \theta_i, s_i^2) p(f(\theta_i) | \theta, \tau^2) p(f(\theta)) p(\tau^2)$$

となり、たとえば $f(\theta)$ の full Bayes 推定量は

$$\hat{\theta}_B = f^{-1} \left(\int f(\theta) \left\{ \int \prod_{i=1}^n p(f(\hat{\theta}_i) | \theta_i, s_i^2) p(f(\theta_i) | \theta, \tau^2) \right. \right. \\ \left. \left. p(f(\theta)) p(\tau^2) df(\theta_1) \dots df(\theta_K) d\tau^2 \right\} df(\theta) \right) \quad (26)$$

$$95\% \text{ CI: } f^{-1} \left(f(\hat{\theta})_B \pm 1.96 \sqrt{\text{Var}^*(f(\hat{\theta})_B)} \right) \quad (27)$$

で与えられる。ここに、上記の積分計算が解析的に評価できないときには MCMC (Markov chain Monte Carlo) 法を利用して数値的に評価するのが簡単である (例、丹後 (2000) 参照)。その際、 $\text{Var}^*(f(\hat{\theta})_B)$ は θ の Markov 連鎖の乱数列の分布 (empirical distribution) の分散として推定する。 $f(\theta_i), \tau^2$ それぞれの full Bayes 推定量も同様に計算できる。

7. Publication bias の問題

統計学の推論の基礎は母集団から必要な大きさの標本を無作為抽出することにある。全数抽出することは実際不可能あるいは時間・費用を節約する観点から無作為抽出を行う。しかし、メタ・アナリシスでは母集団は数えられる程度の有限な数の研究である。したがって、メタ・アナリシスの原則は全数検索である。しかし、同じテーマの研究論文をすべて検索・収集することは下記に解説する様々なバイアスによって不可能に近い。

7.1 統計的に有意なものが優先的に公表される？

特に公表バイアス (publication bias) の問題は深刻である。新薬開発に携わる製薬企業が計画した試験で結果がネガティブであれば公表される可能性は小さいし、研究者にしても思うような結果がでなければ論文を投稿しようとしまいだろう。たとえば、論文を投稿しようとしても雑誌の編集委員会はネガティブな論文は掲載価値が低いと考えて論文採択を否決してしまうかもしれない (最近ではコクラン共同計画の影響でネガティブな結果も掲載する機会は増えているが)。つまり、「公表される、されない」の基準が研究結果の「ポジティブかネガティブ」に強く関連しているからである。したがって、公表された結果だけで (重み付) 平均値を計算すると明らかにポジティブの方向にバイアスがかかってしまう。その典型例としては、癌の化学療法の分野の成績が有名である。進行性卵巣癌患者に対する多剤併用療法とアルキル化薬単独療法との比較試験では、文献サーチで選択された研究に基づく治療効果では多剤併用療法の方が有意に生存率が大きかったのに対し、癌の国際癌データバンク (International Cancer Data Bank) に登録された試験の解析では有意な治療効果の差は観察されていなかったのである (Simes, 1987)。

Published trials ($p=0.0004$)

Registered trials ($p=0.17$)

公表バイアスについて検討したある調査では、1984 年から 1987 年の間に ethics 委員会です承された 285 の臨床試験のうち、1990 年までに約 48% の 138 の結果が公表された。しかし、有意に治療効果があつた研究結果が公表されるオッズは、有意な治療効果が観察されなかった試験結果が公表されるオッズの 3 倍程度高

かった。

また、試験のスポンサーも公表バイアスに大きく関わっている。特に、製薬企業がスポンサーとなって実施された試験は政府・研究者などが中心となって実施された試験より公表される割合は少ない。例えば、上述した進行卵巣癌患者に対する多剤併用療法とアルキル化薬単独療法との比較試験で、多剤併用療法が優れていると発表したのは、製薬企業がスポンサーとなっていた試験では全体の89%であったのに対し、他の臨床試験では61%に過ぎなかった。これは、ネガティブな結果の試験を製薬企業が公表を渋ってきたことを示したデータである。また、多施設共同試験の結果は単一施設で行われた試験の結果より公表される割合は高い。しかし、残念ながら、質の高い研究は質の低い研究より公表される割合は高いかというところでもないようである。

7.2 有意な結果は英語で公表？

これまでの代表的なメタ・アナリシスは主として英語で書かれた臨床試験に基づいている。例えば、1991年から1993年まで実施された26のメタ・アナリシスは英語で書かれた文献に限ったものであった (Gregoire, 1995)。日本のような英語圏でない国の研究者は英語と母国語の両方で論文を書かねばならない点で英語圏の研究者よりハンディがある。ただ、英語が世界の共通語となった現在、いい結果がでたら英語で、さほどでもないければ母国語で、と考える研究者は多いに違いない。したがって、英語の文献だけを収集したメタ・アナリシスにはバイアスがあるといわねばならない。Egger 他 (1996, 1997) はドイツ語で出版された文献について調べた。1985年から1994年までに5つの学術雑誌で公表されたすべてのRCTの論文をmanual searchで探索したのであった。一方で、同じ10年間に英語で公表されたRCTの論文をMEDLINEを利用して検索した。同じファースト著者による英語の論文とドイツ語の論文のペアを集めて比較したのである。英語で公表された論文の63%が有意な結果を示していたがドイツ語で公表された試験結果で有意だったのは35%であった。これらの結果を条件付きロジスティック回帰モデル (meta-regression) で分析したところ、有意な結果を示した試験結果が英語で公表されるオッズはドイツ語のその3.75倍 (95% CI: 1.25-11.3) であった。この関連性は症例数、試験デザイン、試験の質などで調整しても大きな変化が認めら

れなかった (3.98, 95% CI: 1.20-13.2)。症例数、試験デザイン、試験の質などには大きな違いは認められなかった。これらの結果は「英語という言語バイアス」の存在をデータで示したということである。このような状況は日本でも同様であろう。

7.3 引用バイアス

文献の出所を探すとき、上述したコンピュータを利用した検索の限界を補完するために専門家に尋ねたり、他の関連論文の文献リストを利用することが少なくない。しかし、この文献リストを利用するのが実は曲者である。つまり、ここに引用バイアス (citation bias) が潜んでいる。たとえば、ある治療効果を支持した (有意な効果を示した) 論文が支持しない論文より引用される頻度が論文の質、サイズに関わらず高い傾向がつよい。つまり、効果のある論文を選択的に見つけてしまうバイアスが生じるのである。また、論文が公表された論文誌自体が引用バイアスを生んでいることをわすれてはいけない。例えば、New Engl J Medに掲載された論文とほとんど名前も知られていない雑誌とでは引用される頻度は明らかに異なるのであるから…。

7.4 Funnel plots — 公表バイアスの検討

上述したように、必要な文献をすべて、かつ、何の先入観もなく選択することは困難を極める。公表バイアスがある以上、すべて集めることは至難の技であるが、

1. 広義の選択基準を設けてそれを満たす可能なすべての研究を含める
2. 次に、狭義の選択基準を複数設定して5章のべる感度分析を行い、結果のバラツキ、研究間の違いなどを検討する

ことが重要となる。その際、公表バイアスの視覚的な検討ツールとして funnel plots で検討できる。それは

- x 軸にオッズ比、リスク比などの効果の大きさ (effect size) の推定値
- y 軸に例数 (効果の大きさの推定値の精度 = 標準誤差の逆 = 例数に比例)

をプロットしたものである。

1. 全く同じ条件で繰り返し測定すると測定値は真値の回りに対称にばらつき、
2. 例数が大きいとバラツキの大きさが小さくなるので、公表バイアスがなく研究の数が多ければ点のバ

図4. 外科入院患者への心理学的指導が入院期間を減少させるかという研究テーマに関するメタ・アナリシスにおける funnel plots. 上図が公表した論文と未公表の論文を合せた図，下図が公表論文だけの図 (Begg and Berlin, 1988).

ラツキ具合が漏斗(funnel)を逆さにしたような対称形を表す。もっとも、形状が正規分布する保証はどこにもない。図4に外科入院患者への心理学的指導が入院期間を減少させるかという研究テーマに関するメタ・アナリシスにおける funnel plots で、上図が公表した

論文と未公表の論文を合せた図，下図が公表論文だけの図である (Begg and Berlin, 1988)。effect size は標準化された平均値の差である。明らかに下の図は非対称な形となっているのがわかる。図5に β プロックカーのメタ・アナリシスにおける funnel plots を示す。

図5 β ブロッカーのメタ・アナリシスにおける funnel plots.

この分布は少々非対称性を示し、治療効果が認められなかった小規模の試験（右下の裾の部分）が少々欠けていることを示している。funnel plots の対称性の検定の一つである Egger et al. の方法 (1997) を利用すると有意ではない ($p=0.817$)。すでに述べたようにこの場合にはこの公表バイアスの影響は小さかったようである。

最近、公表バイアスへの統計学からの挑戦として、バイアスの検出法、未公表論文数の推定、公表バイアス調整による effect size の推定法などの研究が増加してきている。欠損値の「穴埋め (imputation)」同様、無から有を生み出すのは困難きわまりないが、挑戦は始まったばかりであり今後の新しい進展に期待したい。

最後に、本稿で概説した内容の詳細と方法論の最近の発展については丹後 (2002) にまとめられているので参考になれば幸いである。

参 考 文 献

Beecher, H.K. (1995). The powerful placebo. *JAMA*, **159**, 1602-1606.

Begg, C.B. and Berlin, J.A. (1988). Publication bias: a problem in interpreting medical data. *J R Statist Soc A*, **151**, 419-463.

Cooper, H. and Hedges, L.V. (1994). *The Handbook of Research Synthesis*. Russell Sage Foundation, New York.

DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, **7**, 177-188.

Early breast cancer trialists' collaborative group (1988). Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer. *N Engl J Med*, **319**, 1681-1692.

Egger, M., Smith, G.D., Schneider, M., et al. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, **315**, 629-634.

Gregoire, G., Derderian, F. and Lorient, J.L. (1995). Selecting the language of the publications included in a meta-analysis: Is there a tower of babel bias? *J Clin Epidemiol*, **48**, 159-163.

Glass, G. (1976). Primary, secondary and meta-analysis of research. *Educ Res*, **5**, 3-8.

Gregoire, G., Derderian, F. and Lorient, J.L. (1995). Selecting the language of the publications included in a meta-analysis: Is there a tower of babel bias? *J Clin Epidemiol*, **48**, 159-163.

Lau, J., Antman, E.M., Silva, J.J., et al. (1992). Cumulative meta-analysis of therapeutic trials

- for myocardial infarction. *N Engl J Med*, **327**, 248-254.
- Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *BMJ*, **2**, 1243-1246.
- Simes, R.J. (1987). Confronting publication bias: A cohort design for metaanalysis. *Statistics in Medicine*, **6**, 11-29.
- Yusuf, S., Peto, R., Lewis, J., et al. (1985). Beta blockade during and after myocardial infarction: An overview of the randomized trials. *Progress in Cardiovascular Diseases*, **27**, 335-371.
- 丹後俊郎 (2000). 統計モデル入門. 医学統計シリーズ No. 2, 朝倉書店.
- 丹後俊郎 (2002). メタ・アナリシス—エビデンスの統合をめざす統計手法. 医学統計学シリーズ No. 4, 朝倉書店.