

ベイズ流メタアナリシスにおける外れ値試験の検出と影響力解析

濱口 雄太

総合研究大学院大学 複合科学研究科 統計科学専攻 博士課程(3年次編入学)5年

背景と目的

メタアナリシスでは治療効果の試験間の異質性が大きい事例が存在することがあり、そのような事例では他の試験と比較して極端な値を示す「外れ値」とされる試験が含まれる可能性がある。このような外れ値は、極端な値を示す理由が分からないことが多いものの、明確な理由なく解析対象から除外することはできない。外れ値が存在する場合、変量効果モデルで適切にモデル化できないことが試験間の異質性や平均治療効果の推定に大きく影響を及ぼす可能性がある。これまで、パラメータを追加して変量効果の分布に裾の重い分布や歪んだ分布を仮定し、外れ値の存在を加味したモデルを構築することで平均治療効果の推定に対する影響を低減する手法が提案されている。外れ値が存在する場合は外れ値を解析対象から除外した感度分析が推奨されており、外れ値であると判定する客観的な指標が必要である。古くは、外れ値を検出するために要約統計量をプロットしたグラフによる視覚的な評価が用いられており、主観的な評価であることが課題であった。近年では、回帰分析における影響力を診断する頻度論的な手法が変量効果モデルのメタアナリシスに適用されている。一方で、ベイズ流のメタアナリシスは、一般的に直感的に結果を解釈することが容易であり、平均治療効果や試験間の異質性のパラメータの事前分布を仮定することで過去の試験結果等に基づく事前情報を活用できる長所がある。推定が恣意的になる可能性のある欠点を持ち合わせているものの、頻度論的なメタアナリシスの代替法として挙げられている。

本研究では、ベイズ流の一変量メタアナリシスにおける外れ値試験の検出および影響力解析の手法の特性を事例解析を通じて確認した。

変量効果モデル

一変量メタアナリシスにおける治療効果の試験間の異質性を考慮したモデルとして変量効果モデルが広く用いられている(DerSimonian and Laird, 1986; Whitehead and Whitehead, 1991)。

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

試験数を k 、試験 i の治療効果の推定値を $y_i (i = 1, 2, \dots, k)$ とする。 θ_i は試験 i の治療効果、 μ は平均治療効果、 σ_i^2 は試験内分散、および τ^2 は試験間分散を表す。なお、 σ_i^2 は既知であるとする。治療効果の指標は、一般的に平均差、標準化平均差、リスク差、リスク比、およびオッズ比が用いられ、比の指標は対数変換して用いられる。

外れ値試験の検出方法

外れ値試験の検出についてネットワークメタアナリシス(Zhang et al., 2015)および診断法のメタアナリシス(Matsushima et al., 2020)で提案されていた指標を一変量メタアナリシスに適用した。

相対距離

回帰分析では外れ値を検出するためにleave-one-out交差検証の手法を用いた指標が提案されており、その一つとしてCookの距離に類似した相対距離が挙げられる。一変量メタアナリシスにおける相対距離は以下で定義される。

$$RD_i = \left| \frac{E_{\theta_{new}|y}(\theta_{new}|y) - E_{\theta_{new}|y(i)}(\theta_{new}|y(i))}{E_{\theta_{new}|y}(\theta_{new}|y)} \right|$$

θ_{new} は仮定したモデルにおいて新たな試験を実施した際に想定される治療効果、 $y(i)$ は全ての試験から試験 i を除外したデータを表す。 RD_i は、新たな試験の平均治療効果の事後予測分布における平均が試験 i を除外した場合にどの程度変化するかを表す。本研究では RD_i の閾値を仮定せず、高値である試験を平均治療効果への影響が大きい試験と見なした。

標準化残差

相対距離と同様にleave-one-out交差検証の手法を用いた指標である標準化残差は、ベイズ流の枠組みにおいても広く用いられている。一変量メタアナリシスにおける標準化残差は以下で定義される。

$$SR_i = \frac{y_i - E_{\theta_{new}|y(i)}(\theta_{new}|y(i))}{\sqrt{Var_{\theta_{new}|y(i)}(\theta_{new}|y(i))}}$$

SR_i は、試験 i を除外したデータを使用して算出した新たな試験の治療効果の事後予測分布における平均および分散を使用して算出する。 SR_i はその絶対値が高値であるほど試験 i が外れ値である可能性が高い。閾値は、標準正規分布を参照して上側および下側2.5%分位点とし、 SR_i が1.96より高値または-1.96より低値である場合、試験 i を外れ値と判定した。

ベイズ流P値

事後予測モデルチェックに基づく手法であり、一変量メタアナリシスにおけるベイズ流P値は以下で定義される。

$$P_{D_i} = P[D_i(y_i, \xi) < D_{i,new}(y_{i,new}, \xi) | y]$$

$y_{i,new}$ は仮想的にサンプリングした新たな試験 i の治療効果の推定値、 ξ は全てのパラメータを表す。 P_{D_i} の閾値を一般的な有意水準である5%とする。試験 i の不一致の程度を示す $D_i(y_i, \xi)$ および事後予測分布からサンプリングされた新たな試験 i の不一致の程度を示す

$D_{i,new}(y_{i,new}, \xi)$ は以下で定義される。

$$D_i(y_i, \xi) = \frac{[y_i - E_{\theta_{new}|\xi}(\theta_{new}|\xi)]^2}{Var_{\theta_{new}|\xi}(\theta_{new}|\xi)}, \quad D_{i,new}(y_{i,new}, \xi) = \frac{[y_{i,new} - E_{\theta_{new}|\xi}(\theta_{new}|\xi)]^2}{Var_{\theta_{new}|\xi}(\theta_{new}|\xi)}$$

尺度混合モデルにおける尺度パラメータ

一変量メタアナリシスにおける尺度混合モデルは以下で定義される。

$$\text{model 1} \quad y_i \sim N(\theta_i, \lambda_i \sigma_i^2), \quad \theta_i \sim N(\mu, \tau^2), \quad \lambda_i = 1$$

$$\text{model 2} \quad y_i \sim N(\theta_i, \lambda_i \sigma_i^2), \quad \theta_i \sim N(\mu, \tau^2), \quad \lambda_i \sim \text{Exp}(2)$$

$$\text{model 3} \quad y_i \sim N(\theta_i, \lambda_i \sigma_i^2), \quad \theta_i \sim N(\mu, \tau^2), \quad \lambda_i \sim \text{IG}(1, 1)$$

尺度混合モデルでは試験内分散を試験 i の尺度パラメータである λ_i により調整し、 λ_i が1である場合(model 1)、変量効果モデルと同様に y_i は正規分布に従うが、 λ_i がパラメータ2の指数分布に従う場合(model 2)、 y_i は二重指数分布に従い、 λ_i がパラメータ(1, 1)の逆ガンマ分布に従う場合(model 3)、 y_i は自由度2の t 分布に従う。二重指数分布や t 分布のような正規分布より裾の重い分布を仮定し、それぞれの分布に対する y_i の当てはまりを評価することで外れ値を検出することができる。二重指数分布を仮定したmodel 2および自由度2の t 分布を仮定したmodel 3における λ_i の閾値として、正規分布に従うことを否定することが妥当な状況として λ_i が1以上となる確率が0.95を上回る場合に試験 i を外れ値と判定した。

事例解析

慢性腰痛患者に対する脊椎手技療法(Spinal manipulative therapy: SMT)の有効性を検証したメタアナリシス(Rubinstein et al., 2019)を対象として外れ値試験を検出する4種類の指標の特性を評価した。ベイズ流メタアナリシスの結果を図1に、閾値を設定しない相対距離以外の指標により判定された外れ値試験を除外した場合の結果を表1に示す。

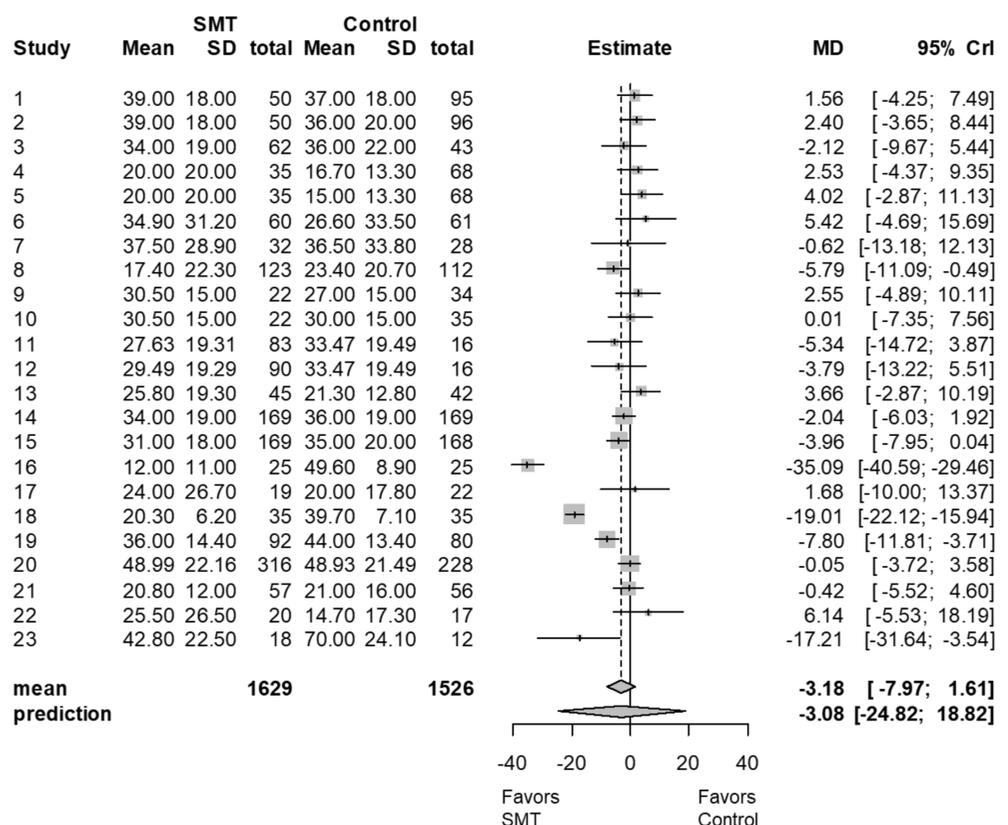


図1 ベイズ流メタアナリシスにおける結果を示したフォレストプロット

表1 全ての研究を用いた結果と、外れ値試験を除外した結果の比較

指標	外れ値と判定された試験	平均治療効果 事後平均[95%信用区間]	試験間分散 事後平均[95%信用区間]
全試験	-	-3.18 [-7.97, 1.61]	114.65 [55.21, 222.00]
標準化残差	16, 23	-1.10 [-4.23, 2.22]	38.79 [16.80, 81.18]
ベイズ流P値	16	-1.62 [-4.91, 1.73]	44.34 [18.25, 94.66]
尺度パラメータ(二重指数分布)	-	-	-
尺度パラメータ(t 分布)	16, 18, 23	-0.53 [-2.47, 1.81]	7.83 [0.04, 26.52]

参考文献

- DerSimonian, R., and Laird, N. M. (1986). *Controlled Clinical Trials* **7**, 177-188.
- Whitehead, A., and Whitehead, J. (1991). *Statistics in Medicine* **10**, 1665-1677.
- Zhang, J., Fu, H., and Carlin, B. P. (2015). *Statistics in Medicine* **34**, 2695-2707.
- Matsushima, Y., Noma, H., Yamada, T., and Furukawa, T. A. (2020). *Research Synthesis Methods* **11**, 237-247.
- Rubinstein, S. M., de Zoete, A., van Middelkoop, M., Assendelft, W. J. J., de Boer, M. R., and van Tulder, M. W. (2019). *BMJ* **364**, 1689.