

診断法のメタアナリシスにおける

極端なプロファイルをもつ研究の検出と影響力の評価

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

背景と課題設定

臨床疫学や医療技術評価の分野において診断法のメタアナリシスの活用が進められており、新しい診断法を評価するための複数の臨床研究が統合された結果が報告されている。診断法のメタアナリシスにおいては、対象となる研究が実施された背景などの相違により、研究間での系統的な異質性が存在し、変量効果モデルを用いて、その異質性をモデル化した解析を行うことが一般的である。しかしながら、実践においては、その異質性の範囲を大きく超えた極端なプロファイルをもつ、外れ値に相当する研究(簡単のため、以降では、外れ値と記載する)が含まれることがあり、その影響力の評価が重要となる。

診断法のメタアナリシスを実施する際に外れ値を特定するための方法に関しては、頻度論に基づく方法が近年提案されたものの[1]、ベイズ流の検出方法についての報告は見当たらない状況である。本研究では、診断法の臨床研究のメタアナリシスに広く適用されている階層ベイズモデルを用いて、外れ値のベイズ流の検出方法を提案する。診断法のメタアナリシスでは、感度および偽陽性率の2変量を同時に評価する必要があるため、2変量およびその相関を統合した指標を提案する。

2変量メタアナリシスモデル

診断法のメタアナリシスの解析にあたって、研究間の異質性をモデル化するために、2変量メタアナリシスを適用し、階層ベイズモデルを用いる。 TP_i, FP_i, FN_i, TN_i を研究 i における真陽性、偽陽性、偽陰性、真陰性の参加者数とする。また、研究 i における陽性または陰性の参加者数を、それぞれ $n_{Ai} = TP_i + FN_i, n_{Bi} = FP_i + TN_i$ とする。

階層ベイズモデルの第1段階目として、 TP_i および FP_i について、以下の2項確率モデルを考える。

$$TP_i \sim \text{Bin}(n_{Ai}, p_{Ai}), FP_i \sim \text{Bin}(n_{Bi}, p_{Bi})$$

ここで、 p_{Ai} および p_{Bi} は感度および偽陽性率である。次に、第2段階目として、ロジット変換された2項確率パラメータ $\theta_{Ai} = \text{logit}(p_{Ai}), \theta_{Bi} = \text{logit}(p_{Bi})$ について、研究間の異質性を考慮するための変量効果モデルを考える。

$$\theta_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_A^2 & \rho\sigma_A\sigma_B \\ \rho\sigma_A\sigma_B & \sigma_B^2 \end{pmatrix}$$

ここで $\theta_i = (\theta_{Ai}, \theta_{Bi})^T, \boldsymbol{\mu} = (\mu_A, \mu_B)^T$ である。階層ベイズモデルにおいて、未知パラメータの事前分布には無情報事前分布を用いる。また、ロジット変換した感度および偽陽性率の推定量を $y_{Ai} = \text{logit}(TP_i/n_{Ai})$ および $y_{Bi} = \text{logit}(FP_i/n_{Bi})$ とする。

極端なプロファイルをもつ研究の検出に関する提案法

Relative Distance

研究 i をメタアナリシスから除外した際の、平均パラメータの推定値の変化の大きさを、その研究を除外しない場合の推定値と比較する指標であり、以下で定義される。

$$RD_i = \frac{\sqrt{(\hat{\eta}_A - \hat{\eta}_{A(i)})^2 + (\hat{\eta}_B - \hat{\eta}_{B(i)})^2}}{\sqrt{\hat{\eta}_A^2 + \hat{\eta}_B^2}}$$

$\hat{\eta}_A = \text{logit}^{-1}(\hat{\mu}_A), \hat{\eta}_B = \text{logit}^{-1}(\hat{\mu}_B)$ は全ての研究から求められる感度、偽陽性率の推定値であり、 $\hat{\eta}_{A(i)}, \hat{\eta}_{B(i)}$ は研究 i を除外して求められる推定値である。Relative distanceが大きい場合、感度、偽陽性率の推定値への影響度が大きい研究と言える。

Standardized Residual

Standardized residualは逸脱度に関する尺度として定義され[2-3]、研究 i の観測値(感度および偽陽性率)と、研究 i を除外したデータセットから得られる事後予測平均の差を、事後予測分散で割った値である。

$$SR_i = \left(\mathbf{y}_i - E_{\mathbf{y}_i|\mathbf{y}_{(i)}}(\mathbf{y}_i|\mathbf{y}_{(i)}) \right)^T V_{\mathbf{y}_i|\mathbf{y}_{(i)}}(\mathbf{y}_i|\mathbf{y}_{(i)})^{-1} \left(\mathbf{y}_i - E_{\mathbf{y}_i|\mathbf{y}_{(i)}}(\mathbf{y}_i|\mathbf{y}_{(i)}) \right)$$

ここで $\mathbf{y}_i = (y_{Ai}, y_{Bi})^T$ である。また、 $\mathbf{y}_{(i)}$ は全体から研究 i のみを除外したデータである。Standardized residualが大きい場合、その研究は、事後予測平均からの乖離が大きい外れ値と考えられる。

Bayesian p-value

Bayesian p-valueは事後予測分布を用いたモデル確認の方法であり、観測データと、階層ベイズモデルから得られる事後予測サンプルの乖離をp値によって評価する方法である[2]。感度および偽陽性率を統合したBayesian p-valueを定義するために、事後予測サンプルからの乖離に関する2次元の尺度を定義する。

$$D_i(\mathbf{y}_i, \boldsymbol{\xi}) = \left(\mathbf{y}_i - E_{\mathbf{y}_i|\boldsymbol{\xi}}(\mathbf{y}_i|\boldsymbol{\xi}) \right)^T V_{\mathbf{y}_i|\boldsymbol{\xi}}(\mathbf{y}_i|\boldsymbol{\xi})^{-1} \left(\mathbf{y}_i - E_{\mathbf{y}_i|\boldsymbol{\xi}}(\mathbf{y}_i|\boldsymbol{\xi}) \right)$$

$\boldsymbol{\xi}$ は、パラメータ全体を表す ($\boldsymbol{\xi} = (\boldsymbol{\mu}^T, \sigma_A^2, \sigma_B^2, \rho)^T$)。次に、事後予測サンプル \mathbf{y}_i^* を用いて、乖離の尺度 $D_i(\mathbf{y}_i^*, \boldsymbol{\xi})$ を同様に定義する。そして、Bayesian p-valueを以下で定義する。

$$P_i = P[D_i(\mathbf{y}_i^*, \boldsymbol{\xi}) > D_i(\mathbf{y}_i, \boldsymbol{\xi}) | \mathbf{y}]$$

このBayesian p-valueが小さい場合、観測値は事後予測サンプルと比較して極端であったことが示され、その研究は外れ値であると考えられる。

SROC曲線のAUCへの影響度

診断法のメタアナリシスにおいて、SROC曲線の曲線下面積(AUC)は診断性能の評価のために広く用いられている。そのため、leave-one-out cross-validationを適用することによってAUCを用いた影響度診断を行うことができる。各研究の影響力は、以下の尺度によって評価される。

$$\Delta AUC_i = \widehat{AUC} - \widehat{AUC}_{(i)}$$

\widehat{AUC} は全ての研究から構成されるデータセットを用いて推定されるAUCであり、 $\widehat{AUC}_{(i)}$ は研究 i を除外したデータセットを用いて推定されるAUCである。

実データへの適用

図1は、小児の尿路感染症患者における膀胱尿管逆流のスクリーニング診断方法として、超音波検査の診断性能を評価したメタアナリシスの結果である[4]。図の白丸は各研究の結果を示し、黒丸はメタアナリシスによる推定値を示す。また、白丸の大きさは参加者数の大小を表す。表1に、提案法の適用結果の要約を示す。提案法で外れ値または影響度の大きな研究を除外して解析したメタアナリシスの結果、感度は0.41から0.44、偽陽性率は0.18から0.23、SROC曲線のAUCは0.565から0.634の範囲の値となった。

提案法を用いて、外れ値や影響力の大きな研究について評価した結果、いくつかの外れ値は感度、偽陽性率、AUCの推定値に影響を与えることが分かった。しかしながら、元々のメタアナリシスの結論を変えるような重大な影響力は認められず、提案法を用いた影響度診断と感度分析の結果、診断法のメタアナリシスの結論の妥当性が確認された。診断法のメタアナリシスにおいては、外れ値または影響力の大きな研究を目視で確認する方法が、これまでは一般的であったが、本研究では客観的、定量的な評価方法を提案した。このような定量的な評価は、診断法のメタアナリシスによるエビデンスの統合に有用であると考えられる。

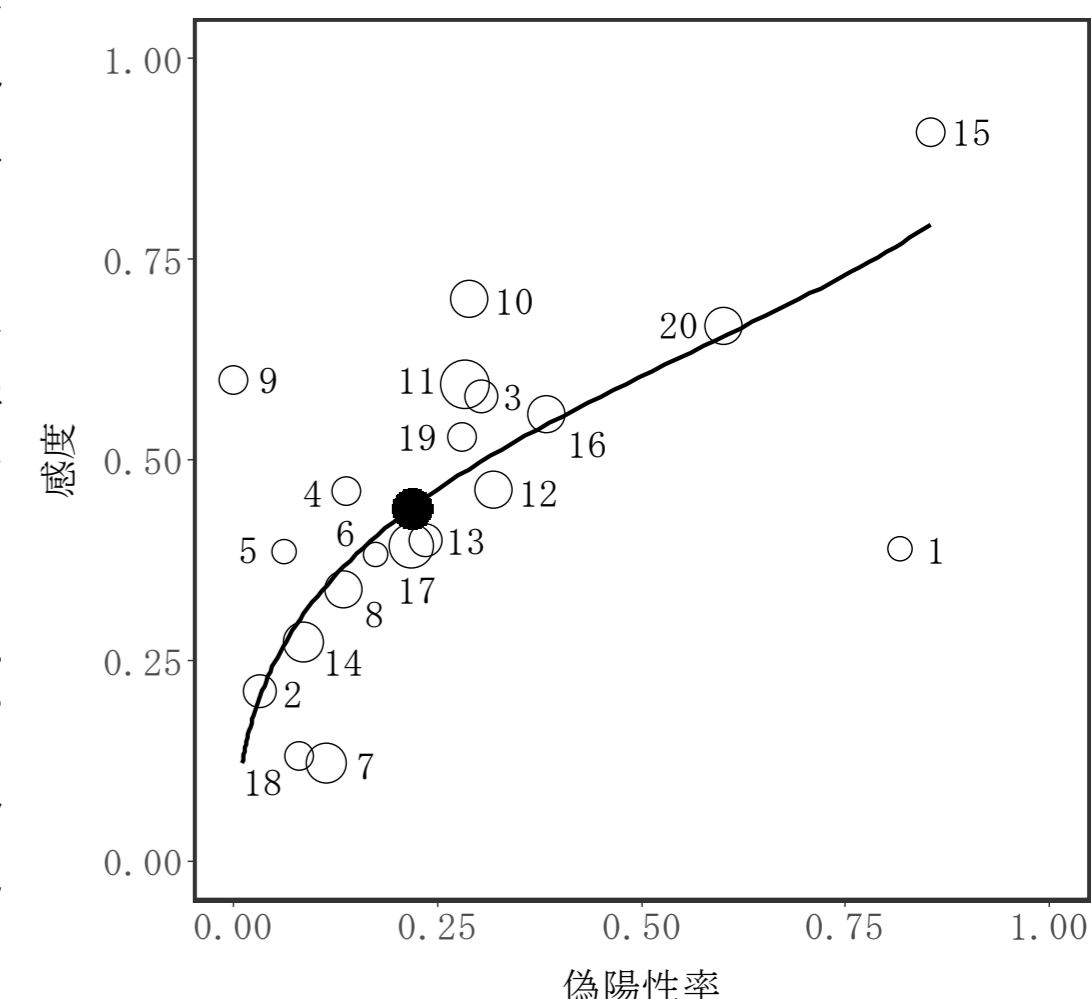


図1 膀胱尿管逆流の超音波による診断のメタアナリシス結果

表1 外れ値とされた研究を除外した場合のメタアナリシスの結果

	外れ値	感度	偽陽性率	SROC曲線のAUC
全ての研究を用いた結果	—	0.44	0.22	0.588
外れ値または影響度の大きな研究を除いた結果				
Relative distance	研究 7, 15	0.44	0.20	0.565
Standardized residual	研究 1, 7, 9, 15	0.43	0.20	0.625
Bayesian p-value	研究 1, 9	0.44	0.23	0.634
SROC曲線のAUCへの影響度	研究 1, 15	0.41	0.18	0.606

- 参考文献 [1] Negeri, Z. F. and Beyene, J. (2020). Statistical methods for detecting outlying and influential studies in meta-analysis of diagnostic test accuracy studies, *Statistical Methods in Medical Research*, **29**(4), 1227-1242.
 [2] Carlin, B. P. and Louis, T. A. (2009). *Bayesian Methods for Data Analysis*, Chapman and Hall/CRC Press, Boca Raton, Florida.
 [3] Zhang, J., Fu, H. and Carlin, B. P. (2015). Detecting outlying trials in network meta-analysis, *Statistics in Medicine*, **34**(19), 2695-2707.
 [4] Shaikh, N., Spingarn, R. B. and Hum, S. W. (2016). Dimercaptosuccinic acid scan or ultrasound in screening for vesicoureteral reflux among children with urinary tract infections, *Cochrane Database of Systematic Reviews*, **7**, Art. No.: CD010657.