

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

博士（統計科学）

総合研究大学院大学
複合科学研究科
統計科学専攻

令和2（2020）年度

概要

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が含まれることがあり、その影響力の評価が重要となる。本論文では、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、外れ値に相当する研究のベイズ流の検出方法を提案する。特に、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。影響力評価のための統合指標として、(1) 相対距離、(2) 標準化残差、(3) ベイズ流 P 値、(4) SROC (summary receiver operating characteristic) 曲線の AUC (area under the curve) への影響度、(5) 尺度混合正規分布における尺度パラメータの事後推定値を提案する。実際のメタアナリシスへの適用として、小児の尿路感染症患者における膀胱尿管逆流の超音波によるスクリーニングを評価したメタアナリシス、および、喘息患者における好酸球性気道炎症の代替マーカーを評価したメタアナリシスに提案手法を適用することによって、提案手法の有用性の評価を行う。

目次

概要	2
目次	3
第 1 章 はじめに	6
第 2 章 診断法のメタアナリシス	9
2.1 2 変量メタアナリシスモデル	9
2.2 外れ値の検出に関する既存手法のレビュー	10
第 3 章 相対距離	16
3.1 背景と目的	16
3.2 感度または偽陽性率に対する相対距離	16
3.3 感度と偽陽性率を同時に評価する相対距離	18
3.4 診断オッズ比を用いた相対距離	19
第 4 章 標準化残差	21
4.1 背景と目的	21
4.2 感度または偽陽性率に対する標準化残差	21
4.3 感度と偽陽性率を同時に評価する標準化残差	22
4.4 診断オッズ比を用いた標準化残差	24
第 5 章 ベイズ流 P 値	25
5.1 背景と目的	25
5.2 感度または偽陽性率に対するベイズ流 P 値	26
5.3 感度と偽陽性率を同時に評価するベイズ流 P 値	27
5.4 診断オッズ比を用いたベイズ流 P 値	28
第 6 章 SROC 曲線の AUC への影響度	29
6.1 背景と目的	29

6.2	SROC 曲線の AUC への影響度	29
第 7 章	尺度混合正規分布における尺度パラメータの事後推定値	31
7.1	背景と目的	31
7.2	感度または偽陽性率に対する尺度パラメータ	32
7.3	感度と偽陽性率を同時に評価する尺度パラメータ	34
第 8 章	実データへの適用（小児の尿路感染症患者における膀胱尿管逆流の超音波によるスクリーニング）	36
8.1	適用データの紹介	36
8.2	計算プログラムと実行条件	40
8.3	提案法の適用結果	41
8.4	メタアナリシスの結果に対する影響力の評価	52
第 9 章	実データへの適用（喘息患者における好酸球性気道炎症の代替マーカー）	58
9.1	適用データの紹介	58
9.2	提案法の適用結果	60
9.3	メタアナリシスの結果に対する影響力の評価	69
第 10 章	考察	75
10.1	研究成果のまとめ	75
10.2	各提案法の特徴	76
10.3	適用上の留意点	78
付録	解析プログラム	81
付録 A	階層ベイズモデルによる診断法のメタアナリシス	81
付録 B	AUC の算出	85
付録 C	相対距離, 標準化残差, SROC 曲線の AUC への影響度	87
付録 D	ベイズ流 P 値	91

付録 E 尺度混合正規分布における尺度パラメータの事後推定値.....	94
謝辞.....	100
参考文献.....	101

第1章 はじめに

臨床疫学や医療技術評価の分野において、診断法の性能評価に関する臨床研究のメタアナリシスは、近年、臨床ガイドラインや医療政策の策定にも広く用いられている重要な研究方法であり、方法論の開発が進められている[1, 2]。診断法のメタアナリシスにおける方法論の発展により、現在では2変量メタアナリシスモデルが、標準的な解析方法の1つとされている。この方法では、診断法の研究からの2つの相関をもつアウトカム（すなわち感度と偽陽性率）を同時にモデル化し、これらの相関関係の情報を統計的推測に組み入れることが可能である。さらに、2変量モデルの解析結果を用いることにより、対応するSROC曲線を推定することも可能である[3-8]。

診断法のメタアナリシスにおいては、対象となる臨床研究が実施された背景、時期、地域、施設、患者集団、診断マーカーのカットオフ値などの相違により、研究間での系統的な異質性が認められることが一般的であるため、変量効果モデルを用いて、異質性をモデル化した解析を行うことが一般的である[4, 8]。しかしながら、実践においては、これら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（以降では簡単のために外れ値と記載する）が含まれることがある。このような外れ値については、メタアナリシスの結果にバイアスを及ぼす可能性があり、最終的な結論に深刻な悪影響を及ぼすおそれもある。そのため、これらの統合解析を行う際には、あらかじめそれぞれの研究が、極端なプロファイルをもつ外れ値となっているか否か、また、その影響力の評価を行うことが重要となる。

診断法のメタアナリシスにおける外れ値の検出方法としては、いくつかの探索的な方法やグラフィカルツールが提案されていたものの[9-12]、それらの方法は解析者の主観に強く依存した評価方法であるという欠点をもっていた[13]。近

年、この問題に対処するために、頻度論に基づく Reitsma の 2 変量の変量効果モデル[4]に適用できる、古典的な回帰診断に基づく方法を、Negeri and Beyene は提案した[13]。彼らの方法は、頻度論の枠組みでの、残差に基づく診断法と、mean-shift outlier model に基づく方法である[13]。一方で、ベイズ流の影響力診断方法も有力なアプローチであり、Carlin and Louis [14]による影響力の診断方法は、外れ値検出のための有用なツールとして、広く用いられている。エビデンス統合の分野においては、Zhang et al. [15]は、Carlin and Louis [14]の影響力診断方法をネットワークメタアナリシスに適用し、ネットワークメタアナリシスでの試験レベルの外れ値を検出するための方法を近年開発した。一方で、診断法のメタアナリシスにおいて階層ベイズモデルは有用で、広く用いられている方法論であるにも関わらず[9, 16]、その枠組みにおいて、外れ値の評価、および、影響力解析を行うための方法は、診断法のメタアナリシスに対しては確立されていない状況である。

本論文においては、診断法のメタアナリシスで外れ値を検出する方法、および、その影響度診断の方法について提案を行う。診断法のメタアナリシスで各研究の影響度を評価する際には、感度、偽陽性率およびその相関関係を同時に評価する必要がある。そのため、本論文では、2 変量の変量効果モデルの解析結果を用いて、2 変量の影響度を統合して評価する指標を開発する。外れ値、または、影響度の大きな研究を検出するためのベイズ流の方法として、本論文では、以下の 5 つの方法を提案する: (1) 相対距離, (2) 標準化残差, (3) ベイズ流 P 値, (4) SROC 曲線の AUC への影響度, (5) 尺度混合正規分布 (scale mixtures of normals) における尺度パラメータの事後推定値。なお、外れ値と影響度の大きな研究は異なるものであり、例えば、参加者数の多い研究は外れ値でなくても、影響度の大きな研究となりうる。本論文の提案法は、極端なプロファイルをもつ外れ値や、全体の統合解析への影響度の大きな研究を発見することを目的としている。提

案法の性能は，実データへの適用を通して確認する。

本論文は以下のように構成される。2章では，診断法のメタアナリシスで用いられる2変量メタアナリシスについて概略を述べ，外れ値の検出に関する既存手法をレビューする。3章から7章では外れ値を検出し，影響力診断を行うための5つの提案法を示す。8章と9章では，診断法のメタアナリシスについての2つの事例（小児の尿路感染症患者における膀胱尿管逆流の超音波によるスクリーニング[17]，および，喘息患者における好酸球性気道炎症の代替マーカー[18]）への適用を通して，提案法の有用性を示す。最後に，10章が考察となる。なお，本論文の3章から6章の提案法，および，8章の実データへの適用については，原著論文として公表されている（Matsushima et al. [19]）。

第2章 診断法のメタアナリシス

2.1 2変量メタアナリシスモデル

本論文では、診断法のメタアナリシスの解析にあたり、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こととする[9,16]。表2.1のように、 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 ($i = 1, 2, \dots, N$)。また、研究 i において、真の有病状態が疾患あり、または、疾患なしとなる参加者数を、それぞれ $n_{Ai} = TP_i + FN_i$, $n_{Bi} = FP_i + TN_i$ とする。

表 2.1 診断法の研究 i における診断結果 (2×2 分割表)

診断結果	真の有病状態	
	疾患	非疾患
疾患	真陽性 TP_i	偽陽性 FP_i
非疾患	偽陰性 FN_i	真陰性 TN_i
合計	n_{Ai}	n_{Bi}

そして、ロジット変換 ($\text{logit}(x) = \log\{x/(1-x)\}$) 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。

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。

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

ここで、 p_{Ai} および p_{Bi} は研究 i における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 $\theta_i = (\theta_{Ai}, \theta_{Bi})^T$, $\boldsymbol{\mu} = (\mu_A, \mu_B)^T$ であり、 μ_A および μ_B はロジット変換された感度および偽陽性率に関する母平均パラメータである。 σ_A^2 および σ_B^2 は、それぞれ θ_{Ai} および θ_{Bi} の異質性に関する分散パラメータであり、 ρ は相関パラメータである。なお、 TP_i および FP_i は、変量効果パラメータ θ_{Ai} および θ_{Bi} が与えられた下で、条件付き独立となる。

ベイズ推定における事前分布としては、一般的に用いられるような無情報事前分布を用いることとし、具体的には $\mu_A, \mu_B \sim N(0, 100)$, $\sigma_A, \sigma_B \sim U(0, 10)$, $\rho \sim U(-1, 1)$ とする。モデルパラメータの事後分布は、マルコフ連鎖モンテカルロ法 (Markov Chain Monte Carlo, MCMC) [20, 21]により推定することができる。

本節についての解析プログラムは、付録Aにまとめる。

2.2 外れ値の検出に関する既存手法のレビュー

本節では、診断法のメタアナリシスにおける外れ値の検出に関する既存研究について述べる。診断法のメタアナリシスで外れ値を調べる方法としては、フォレストプロット、または、SROC 曲線とともに描かれる感度と偽陽性率の散布図を用いて、目視で確認する方法がこれまでは一般的であった。しかしながら、既存研究は多くないものの、上記以外のプロットや統計的な外れ値の検出方法も提案されているので、本節で紹介する。

Verde [9, 10]による方法

Verde [9, 10]は、研究間の異質性が大きい状況においても、モデルパラメータについての頑健な推定値を得ることを目的として、研究間の異質性に関する分散共分散行列に、研究ごとの重みを付与し外れ値の影響を軽減させる方法を提案した。以下で、Verde [9, 10]の提案法を紹介する。

Verde [9, 10]が提案する階層ベイズモデルでは、まず 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。

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

次に、

$$D_i = g(p_{Ai}) - g(p_{Bi}), \quad S_i = g(p_{Ai}) + g(p_{Bi})$$

とおく。 $g(\cdot)$ はリンク関数であり、ロジット関数などがデータに応じて選択される。 D_i は診断法の識別性能に関連する値であり、 D_i が大きいほど診断性能が良くなる。一方で S_i は診断法の閾値に関連する値であり、 S_i が大きい場合は、感度および偽陽性率がともに高くなる。 D_i および S_i は以下のようにモデルされる。

$$(D_i, S_i) \sim N(\boldsymbol{\mu}, \boldsymbol{\Psi}_i), \quad i = 1, 2, \dots, N,$$

$$\boldsymbol{\Psi}_i = w_i \times \boldsymbol{\Lambda},$$

$$w_i \sim p(w_i)$$

$\boldsymbol{\Lambda}$ は精度に関する行列（共分散行列の逆行列）であり、 $p(w_i)$ は尺度パラメータの確率密度である。 $p(w_i)$ によって、 (D_i, S_i) の周辺分布のモデルに柔軟性を与えることができる。具体的には、 w_i の事前分布をパラメータ $(v/2, v/2)$ のガンマ分布 ($w_i \sim \Gamma(v/2, v/2), v > 2$) とすることにより、誤差分布を2変量t分布に拡張することができる。また、 w_i の事前分布をパラメータ1の指数分布 ($w_i \sim \text{Exp}(1)$) とすることにより、誤差分布を二重指数分布に拡張することができる。 w_i の事後分布は、研究間の極端な異質性を検出するために用いることが

でき、異質性の大きな研究については、 w_i が 1 よりも相当小さな値（例えば 0.7 よりも小さい値）をとるとされている。上記の手法は、R の `bamdit` パッケージに実装されている[10]。

Stata でのプロット

統計解析ソフトウェアの Stata の `midas` モジュールは診断法のメタアナリシスのためのモジュールであり、診断法のメタアナリシスの解析を行い、様々なグラフを作成する機能を備えている。そして、グラフの中には、診断法のメタアナリシスにおける外れ値の発見を助けるグラフが含まれている[11]。`midas` モジュールでは、まず、2 変量モデルによって診断法のメタアナリシスが実行され、その後、グラフによってモデルへの当てはまりを確認することができる。具体的には、残差に関する正規確率プロットにより誤差分布の正規性の確認ができ、Cook の距離を用いたプロットにより影響度の大きな研究を確認することができる。また、標準化残差の散布図、および、ロジット変換された感度および特異度の 2 変量箱ヒゲ図が用意されており、外れ値を目視で探すことができる。

同じく統計解析ソフトウェアの Stata の `metandi` モジュール[12]も診断法のメタアナリシスのためのモジュールであり、`midas` モジュールと同様に、診断法のメタアナリシスの解析を行って、種々のグラフを作成する機能を備えている。そして、診断法のメタアナリシスにおける外れ値の発見をうながすグラフが用意されている。`metandi` モジュールでは、まず、2 変量モデルおよび HSROC (hierarchical SROC) モデルによって診断法のメタアナリシスが実行され[4, 5]、各研究の感度および特異度の事後予測値が求められる。その後、Cook の距離による影響度のグラフ、および、標準化残差の散布図を出力することができる。なお、Cook の距離による影響度のグラフは、分散パラメータも含む全てのモデルパラメータへの影響を一括で評価する方法が採用されている。

Negeri and Beyene [13]による方法

Negeri and Beyene [13]は、頻度論に基づく診断法のメタアナリシスについて、外れ値および影響度の大きな研究の、古典的な回帰診断に基づく検出方法を提案した。Negeri and Beyene [13]の研究は、診断法のメタアナリシスにおける外れ値の検出を明確に志向した研究である。以下で、Negeri and Beyene [13]の提案法を紹介する。

Negeri and Beyene [13]の方法では、Reitsma et al. [4]の2変量変量効果モデルを用いる。具体的には、研究 i の感度および特異度を Se_i および Sp_i として、 $\mathbf{y}_i = (y_{1i}, y_{2i})^T = (\text{logit}(Se_i), \text{logit}(Sp_i))^T$ とする。次に、研究内のモデルとして、

$$\mathbf{y}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Psi}_i), i = 1, \dots, N$$

とする。ここで $\boldsymbol{\Psi}_i$ は研究内の共分散行列を表す。次に、研究間のモデルとして、

$$\boldsymbol{\mu}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}), i = 1, \dots, N$$

とする。 $\boldsymbol{\Sigma}$ は研究間の異質性に関する共分散行列である。

Negeri and Beyene [13]の外れ値の検出方法の1つ目は残差に基づく方法であり、 i 番目の研究をメタアナリシスから除外して求められるスチューデント化残差は以下のように定義される。

$$t_i = \left(\text{Var}(\epsilon_{i(-i)}) \right)^{-\frac{1}{2}} \epsilon_{i(-i)} = \frac{y_i - \hat{\mu}_{(-i)}}{\sqrt{\text{Var}(y_i - \hat{\mu}_{(-i)})}}$$

y_i はロジット変換された感度または特異度であり、 $\hat{\mu}_{(-i)}$ は研究 i がメタアナリシスから除外された場合の平均パラメータ（感度または特異度）の推定値である。

Negeri and Beyene [13]は、 t_i が標準正規分布に従うことを示し、外れ値を検出するための閾値として、標準正規分布の上側/下側 2.5%点である ± 1.96 を提案した。さらに、より適切な方法としてパラメトリックブートストラップ法により標本分布の推定を行い、閾値を設定する方法を提案した。

Negeri and Beyene [13]の外れ値の検出方法の2つ目は mean-shift outlier model を用いる方法であり、研究間の異質性について以下のようにモデル化する。

$$\boldsymbol{\mu}_i \sim N(\boldsymbol{\mu} + \boldsymbol{\eta}_j, \boldsymbol{\Sigma}), \quad \boldsymbol{\eta}_j = \begin{cases} (\eta_1, \eta_2)^T, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

上記のモデルを用いることにより $\boldsymbol{\eta}_j$ が推定され、 $\hat{\boldsymbol{\eta}}_j$ の絶対値が大きい場合に外れ値の可能性が示唆される。さらに、以下の帰無仮説および対立仮説に対する尤度比検定を導入した。

$$H_0: \boldsymbol{\eta}_j = 0 \text{ vs } H_1: \boldsymbol{\eta}_j \neq 0$$

検定統計量は、 $LRT_j = -2(l(\boldsymbol{\mu}, \boldsymbol{\Sigma}) - l(\boldsymbol{\mu}, \boldsymbol{\eta}_j, \boldsymbol{\Sigma}))$ である。 $l(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ は帰無仮説の下での対数尤度であり、 $l(\boldsymbol{\mu}, \boldsymbol{\eta}_j, \boldsymbol{\Sigma})$ は対立仮説の下での対数尤度である。帰無仮説が正しい場合に LRT_j は漸近的に自由度2のカイ2乗分布に従うことを示し、外れ値を検出するための閾値を提案した。さらに、より適切な方法としてパラメトリックブートストラップにより標本分布の推定を行い、閾値を設定する方法を提案した。

さらに Negeri and Beyene [13]は、影響度の大きな研究の検出方法として、以下の2つの指標を提案した。1つ目の指標は以下で定義される。

$$\text{VARCOVRATIO}_i = \frac{\det(\text{VARCOV}_{(-i)})}{\det(\text{VARCOV})} = \frac{\det((\boldsymbol{\Sigma} \widehat{\boldsymbol{W}}_{i(-i)})^{-1})}{\det((\boldsymbol{\Sigma} \widehat{\boldsymbol{W}}_i)^{-1})}$$

ここで $\widehat{\boldsymbol{W}}_i = (\widehat{\boldsymbol{\Sigma}} + \boldsymbol{\Psi}_i)^{-1}$ であり、 $\widehat{\boldsymbol{W}}_{i(-i)}$ は研究 i を除いて得られる $\widehat{\boldsymbol{W}}_i$ の値である。研究 i の影響力が大きい場合、VARCOVRATIO の値は0に近くなる。また、もう1つの指標は以下で定義される。

$$\text{SIGMARATIO}_i = \frac{\det(\widehat{\boldsymbol{\Sigma}}_{(-i)})}{\det(\widehat{\boldsymbol{\Sigma}})}$$

$\widehat{\boldsymbol{\Sigma}}$ は $\boldsymbol{\Sigma}$ の推定量であり、 $\widehat{\boldsymbol{\Sigma}}_{(-i)}$ は研究 i を除いて得られる推定量である。研究 i の影響力が大きい場合、SIGMARATIO の値は0に近くなる。

既存研究に関するまとめ

以上で、診断法のメタアナリシスにおける外れ値の検出に関する既存研究について述べた。

Stata の `midas` および `metandi` モジュールは[11, 12], 診断法のメタアナリシスのためのモジュールであり, 機能の 1 つとして外れ値の検出を助けるグラフを備えている。しかしながら, これらは外れ値の検出と影響力の評価を目的とした研究ではなく, プロットの目視による外れ値の検出は, 解析者の主観に依存するという欠点がある。

Verde の方法[9, 10]は外れ値の検出と重みづけを含む提案であるものの, 外れ値の検出と影響力の評価に主眼をおいたものではない。また, Verde [9, 10]は, 研究間の異質性分散に, 研究ごとに異なる尺度パラメータを導入することを提案したが, Carlin and Louis [14]による影響力の診断方法には, 他の診断方法も存在する。

一方で, Negeri and Beyene [13]の報告は, 外れ値の検出と影響力の評価について, 事例への適用も含めて詳細に議論されているものの, 頻度論に基づく外れ値の検出方法である。

従って, 診断法のメタアナリシスにおいて階層ベイズモデルは有用であり頻用されている方法論であるにも関わらず[9, 16], Carlin and Louis [14]による, ベイズ流の影響力診断方法は診断法のメタアナリシスに対しては確立されていない状況であった。本論文においては, ベイズ流の方法を用いて, 外れ値の検出および影響力診断について, 診断法のメタアナリシスに適した検出方法を提案する。

第3章 相対距離

3.1 背景と目的

1つ目の提案手法は、相対距離 (relative distance, RD) である。この提案法は Zhang et al. [15]がネットワークメタアナリシスへの適用のために開発した指標である。相対距離は、 i 番目の研究をメタアナリシスから除外する leave-one-out 交差検証法によって、平均パラメータの推定値への影響を評価する指標である。

本章では、最初に、相対距離を診断法のメタアナリシスに適用することにより、感度または偽陽性率のどちらか一方のみを評価する 1 変量の相対距離を提案する。次に、感度と偽陽性率を同時に評価する相対距離として、Zhang et al. [15]によって提案された平均相対距離を診断法のメタアナリシスに適用するとともに、新しい指標である統合相対距離を提案する。最後に、診断法のメタアナリシスのための要約指標を用いる方法として、診断オッズ比を用いる相対距離を提案する。

3.2 感度または偽陽性率に対する相対距離

本節では、感度または偽陽性率のどちらか一方のみを評価する 1 変量の相対距離について提案する。研究 i の感度および偽陽性率についての周辺的な相対距離は、Zhang et al. [15]と同様に定義することができ、次の式で得られる。

$$RD_{Ai} = \frac{\hat{\eta}_A - \hat{\eta}_{A(i)}}{\hat{\eta}_A}, RD_{Bi} = \frac{\hat{\eta}_B - \hat{\eta}_{B(i)}}{\hat{\eta}_B}$$

ここで $\hat{\eta}_A = \text{logit}^{-1}(\hat{\mu}_A)$, $\hat{\eta}_{A(i)} = \text{logit}^{-1}(\hat{\mu}_{A(i)})$, $\hat{\eta}_B = \text{logit}^{-1}(\hat{\mu}_B)$, $\hat{\eta}_{B(i)} = \text{logit}^{-1}(\hat{\mu}_{B(i)})$ である。 $\hat{\mu}_A$ および $\hat{\mu}_B$ は、それぞれ μ_A および μ_B の推定量である。

り、全ての研究を解析に含めたメタアナリシスにより推定される。一方で $\hat{\mu}_{A(i)}$ および $\hat{\mu}_{B(i)}$ は、研究 i を除く全ての研究を含めたメタアナリシスにより推定される。これらの推定量は 2.1 節の階層ベイズモデルにより求めることができ、通常は、事後平均が推定量として用いられる。絶対値として、大きな相対距離をもつ研究は、感度および偽陽性率についての周辺の観点で、影響力の大きい研究と判断される。

提案法においては、評価指標は平均パラメータ $\hat{\mu}_A$ および $\hat{\mu}_B$ でなく、0~1 のスケールに逆ロジット変換した $\hat{\eta}_A$ および $\hat{\eta}_B$ で定義している。逆ロジット変換しないスケールの $\hat{\mu}_A$ および $\hat{\mu}_B$ を用いると、感度または偽陽性率の全体の推定値が 50% の場合に相対距離の分母が 0 となり、相対距離が算出不能となる。この点を回避するために、相対距離の定義においては、逆ロジット変換した推定値 $\hat{\eta}_A$ および $\hat{\eta}_B$ を用いることとする。なお、感度および偽陽性率が 0 に近い場合、相対距離の分母 $\hat{\eta}_A$ または $\hat{\eta}_B$ も 0 に近い値となり、その結果、相対距離は大きい値になる。逆に、 $\hat{\eta}_A$ または $\hat{\eta}_B$ が 1 に近い場合には、影響度の大きな研究が存在したとしても、相対距離は小さな値となるかもしれない。

Zhang et al. [15] での相対距離は、本節での相対距離とは異なり、絶対値をとることによって、符号の情報を消した形である。しかしながら、感度や偽陽性率に対して、正の方向への影響力をもつか、負の方向への影響力をもつかは重要な情報であるので、本節では、絶対値をとらない形とした。

なお、診断法のメタアナリシスにおいては、陽性と陰性を判定するためのカットオフ値は研究ごとに異なる。そして、各研究では、それぞれの研究におけるカットオフ値の条件のもとで、感度と偽陽性率の推定を行うため、各研究での推定対象は異なっていると考えることができる。leave-one-out 交差検証法に基づく相対距離は、このような各研究での推定対象の違いやカットオフ値の違いが、全体の推定値に与える影響を評価していると考えられる。例えば、相対距離を用いる

ことで、感度と偽陽性率の推定値が、一部の、極端なカットオフ値をもつ研究の影響を強く受けていないかどうかを確認することができる。

3.3 感度と偽陽性率を同時に評価する相対距離

本節では、感度および偽陽性率を同時に評価する指標について提案する。診断法のメタアナリシスにおいては、推定の主な目的は、平均パラメータ μ を推定することであり、 μ は感度、偽陽性率およびそれらの相関に関する情報から推定される。従って影響力についても2次元の指標で評価することが適切である。

最初に、Zhang et al. [15]によって提案された平均相対距離（average relative distance, ARD）を診断法のメタアナリシスに適用する。平均相対距離は、感度に対する相対距離 RD_{Ai} と、偽陽性率に対する相対距離 RD_{Bi} の平均値で定義され、次の式で書き表せる。

$$ARD_i = \frac{1}{2} \left(\left| \frac{\hat{\eta}_A - \hat{\eta}_{A(i)}}{\hat{\eta}_A} \right| + \left| \frac{\hat{\eta}_B - \hat{\eta}_{B(i)}}{\hat{\eta}_B} \right| \right)$$

この、平均相対距離は2つの1次元のユークリッド距離を平均した指標である。平均相対距離の定義においては、感度や偽陽性率に対する相対距離とは異なり、絶対値をとる形としている。理由としては、感度に対しては正の方向の影響、偽陽性率に対しては負の方向の影響となった場合などに、感度と偽陽性率に対する影響が相殺されることを防ぐためである。

次に、本論文では、感度と偽陽性率を同時に評価する以下の指標を提案し、これを統合相対距離（synthetic relative distance, SRD）と呼ぶこととする。

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

統合相対距離は、2次元空間での相対的なユークリッド距離として定義されてお

り、全ての研究を用いて得られた推定値と、ある研究を除いたデータから得られた推定値の差を定量的に評価するものである。診断法のメタアナリシスにおいては、感度と偽陽性率の散布図が頻用される。このような 2 次元空間上での影響力評価においては、1 次元のユークリッド距離を平均した平均相対距離よりも、2 次元空間での相対的なユークリッド距離である統合相対距離の方が自然であると考えられる。平均相対距離と統合相対距離の数値的な比較については、実データへの適用結果（8.3 節，9.2 節）で言及する。

なお、平均相対距離や統合相対距離は、感度と偽陽性率への影響度を同時に評価するため、感度や偽陽性率のそれぞれに対する影響度の大きさや、影響度の方向は分かりにくくなる。そのため、感度に対する相対距離、または、偽陽性率に対する相対距離の方が、影響力に関する臨床的な解釈は容易である。

一方で、本論文では、診断法のメタアナリシスに対して、2 変量メタアナリシスモデルを用いている。従って、推定結果への影響力の評価においても、1 変量ずつの評価のみではなく、2 変量を同時に評価することが重要であり、1 変量の周辺的な相対距離と、2 変量を同時に評価する相対距離はともに用いられるべきと考える。

3.4 診断オッズ比を用いた相対距離

相対距離に関する最後の提案法として、診断法のメタアナリシスのための要約指標を用いる方法を提案する。診断法の性能を評価するためのよく知られた要約指標として、診断オッズ比 (diagnostic odds ratio, DOR) [22]がある。診断法の評価においては感度および偽陽性率の 2 つの観点から性能を評価する必要があるが、診断オッズ比はそれを 1 次元に集約している点で有用である。診断オッズ比は、疾病ありの集団における陽性のオッズ $p_{Ai}/(1-p_{Ai})$ と、疾病なしの

集団における陽性のオッズ $p_{Bi}/(1-p_{Bi})$ の比として定義される。

$$DOR = \frac{p_{Ai}/(1-p_{Ai})}{p_{Bi}/(1-p_{Bi})}$$

診断性能が高い（疾病ありの集団では陽性のオッズが高く，疾病なしの集団では陽性のオッズが低い）ほど，診断オッズ比は大きな値をとる。なお，診断オッズ比と陽性尤度比（ $LR_{(+)}$ ）および陰性尤度比（ $LR_{(-)}$ ）の間には次の関係が成立する。

$$DOR = \frac{p_{Ai}/(1-p_{Ai})}{p_{Bi}/(1-p_{Bi})} = \frac{p_{Ai}/p_{Bi}}{(1-p_{Ai})/(1-p_{Bi})} = \frac{LR_{(+)}}{LR_{(-)}}$$

感度および偽陽性率への影響を同時に評価するために，診断オッズ比を用いた相対距離を，以下で定義する。

$$RD_{DORi} = \frac{\widehat{DOR} - \widehat{DOR}_{(i)}}{\widehat{DOR}}$$

ここで， \widehat{DOR} はすべての研究から推定される診断オッズ比であり，

$$\begin{aligned} \widehat{DOR} &= \frac{\hat{\eta}_A/(1-\hat{\eta}_A)}{\hat{\eta}_B/(1-\hat{\eta}_B)} = \exp\left(\log\frac{\frac{\hat{\eta}_A}{1-\hat{\eta}_A}}{\frac{\hat{\eta}_B}{1-\hat{\eta}_B}}\right) = \exp(\text{logit}(\hat{\eta}_A) - \text{logit}(\hat{\eta}_B)) \\ &= \exp(\hat{\mu}_A - \hat{\mu}_B) \end{aligned}$$

により得られる。同様に， $\widehat{DOR}_{(i)}$ は研究 i を除外して推定される診断オッズ比であり， $\widehat{DOR}_{(i)} = \exp(\hat{\mu}_{A(i)} - \hat{\mu}_{B(i)})$ により得られる。

3.2節の感度または偽陽性率に対する相対距離と同様であるが，Zhang et al. [15]での相対距離と異なり，本節の診断オッズ比を用いた相対距離の定義では，絶対値をとらない形とし，研究 i が正の方向への影響力をもつか，負の方向への影響力をもつかが分かるようにした。

3章についての解析プログラムは，付録Cにまとめる。

第4章 標準化残差

4.1 背景と目的

2つ目の提案手法は、標準化残差 (standardized residual, SR) である。標準化残差は、Carlin and Louis [14]によって解析モデル診断の指標として述べられており、leave-one-out 交差検証法に基づく方法である。標準化残差は逸脱度に関する指標として定義され、研究 i の観測値と、研究 i が除外されたデータセット (leave-one-out データセット) から得られる事後予測平均の差を、事後予測分散の平方根で割った値である。Zhang et al. [15]は、ネットワークメタアナリシスに対する単変量の影響力診断指標として標準化残差を提案した。

本章では、最初に、標準化残差を診断法のメタアナリシスに適用することにより、感度または偽陽性率のどちらか一方のみを評価する1変量の標準化残差を提案する。次に、感度と偽陽性率を同時に評価する標準化残差として、Zhang et al. [15]によって提案された平均標準化残差を診断法のメタアナリシスに適用するとともに、2変量への拡張を行った統合標準化残差を提案する。最後に、診断法のメタアナリシスのための要約指標を用いる方法として、診断オッズ比を用いる標準化残差を提案する。

4.2 感度または偽陽性率に対する標準化残差

本節では、感度または偽陽性率のどちらか一方のみを評価する標準化残差について提案する。

研究 i の感度および偽陽性率についての周辺的な標準化残差は、Zhang et al. [15]と同様に定義することができ、次の式で得られる。

$$SR_{Ai} = \frac{y_{Ai} - E_{y_{Ai}|y^{(i)}}(y_{Ai}|\mathbf{y}^{(i)})}{\sqrt{\text{Var}_{y_{Ai}|y^{(i)}}(y_{Ai}|\mathbf{y}^{(i)})}}, SR_{Bi} = \frac{y_{Bi} - E_{y_{Bi}|y^{(i)}}(y_{Bi}|\mathbf{y}^{(i)})}{\sqrt{\text{Var}_{y_{Bi}|y^{(i)}}(y_{Bi}|\mathbf{y}^{(i)})}}$$

ここで、 $E_{y_{Ai}|y^{(i)}}(y_{Ai}|\mathbf{y}^{(i)})$ および $E_{y_{Bi}|y^{(i)}}(y_{Bi}|\mathbf{y}^{(i)})$ は、感度および偽陽性率に関する事後予測平均であり、 $\text{Var}_{y_{Ai}|y^{(i)}}(y_{Ai}|\mathbf{y}^{(i)})$ および $\text{Var}_{y_{Bi}|y^{(i)}}(y_{Bi}|\mathbf{y}^{(i)})$ は事後予測分散である。ここでの事後予測平均および事後予測分散は、研究 i を除いたデータセットから求められる。 SR_{Ai} または SR_{Bi} が大きな値となる場合に、対応する研究が外れ値と判断される。これらの SR_{Ai} および SR_{Bi} は周辺的な情報のみを反映した指標である。

標準化残差における事後予測分散については、研究内分散および研究間分散の和として定義される。異質性に関する研究間分散は MCMC より推定することができるものの、研究 i は leave-one-out 交差検証法によって計算から除かれているため、研究 i の研究内分散については MCMC から推定することはできない。そのため、研究内分散 (感度に対しては ω_{iA}^2 , 偽陽性率に対しては ω_{iB}^2 とする) は、デルタ法 ($\text{Var}(f(x)) \approx [f'(x)]^2 \cdot \text{Var}(x)$) を用いることにより、以下の式より推定する[4]。

$$\omega_{iA}^2 = \frac{1}{(TP_i + FN_i) \cdot \frac{TP_i}{(TP_i + FN_i)} \cdot \frac{FN_i}{(TP_i + FN_i)}},$$

$$\omega_{iB}^2 = \frac{1}{(TN_i + FP_i) \cdot \frac{TN_i}{(TN_i + FP_i)} \cdot \frac{FP_i}{(TN_i + FP_i)}}$$

4.3 感度と偽陽性率を同時に評価する標準化残差

本節では、感度および偽陽性率を同時に評価する標準化残差について提案する。まず、Zhang et al. [15]によって提案された平均標準化残差 (average standardized residual, ASR) を診断法のメタアナリシスに適用することにより、感度と偽陽性

率を統合するための別の指標を作成することができる。

$$ASR_i = \frac{1}{2} \left(\left| \frac{y_{Ai} - E_{y_{Ai}|y_{(i)}}(y_{Ai}|\mathbf{y}_{(i)})}{\sqrt{Var_{y_{Ai}|y_{(i)}}(y_{Ai}|\mathbf{y}_{(i)})}} \right| + \left| \frac{y_{Bi} - E_{y_{Bi}|y_{(i)}}(y_{Bi}|\mathbf{y}_{(i)})}{\sqrt{Var_{y_{Bi}|y_{(i)}}(y_{Bi}|\mathbf{y}_{(i)})}} \right| \right)$$

平均相対距離と同様であるが、平均標準化残差の定義においても、絶対値を用いている。感度に対しては正の方向の残差、偽陽性率に対しては負の方向の残差となった場合など、感度と偽陽性率に対する残差が相殺されることを防ぐためである。平均標準化残差は、感度と偽陽性率の両面から、標準化残差を測る指標となっているものの、感度と偽陽性率の間の相関が考慮されていない点で不十分な指標であると考えられる。

従って、本研究では、感度、偽陽性率およびその相関を考慮した指標について提案する。そして、その指標を統合標準化残差 (synthetic standardized residual, SSR) と名付けることとする。

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

ここで、 $\mathbf{y}_i = (y_{Ai}, y_{Bi})^T$ 、 $E_{\mathbf{y}_i|\mathbf{y}_{(i)}}(\mathbf{y}_i|\mathbf{y}_{(i)}) = \left(E_{y_{Ai}|y_{(i)}}(y_{Ai}|\mathbf{y}_{(i)}), E_{y_{Bi}|y_{(i)}}(y_{Bi}|\mathbf{y}_{(i)}) \right)^T$ であり、 $V_{\mathbf{y}_i|\mathbf{y}_{(i)}}(\mathbf{y}_i|\mathbf{y}_{(i)})$ は leave-one-out データセットから得られる事後予測共分散行列である。統合標準化残差は 2 次元空間における標準化残差として定義することができるが、研究 i の観測値についての、leave-one-out データセットからの逸脱度を定量化する指標である。統合標準化残差は、分母に 2 変量の共分散に関する情報を含む点で、平均標準化残差より適切な指標であると考えられる。平均標準化残差と統合標準化残差の数値的な比較については、実データへの適用結果 (8.3 節, 9.2 節) で言及する。

4.4 診断オッズ比を用いた標準化残差

標準化残差に関する最後の提案法として、診断オッズ比を用いることによる標準化残差を提案する。

$$SR_{DORi} = \frac{\log \widehat{DOR}_i - E_{\log DOR_i | \mathbf{y}^{(i)}}(\log DOR_i | \mathbf{y}^{(i)})}{\sqrt{\text{Var}_{\log DOR_i | \mathbf{y}^{(i)}}(\log DOR_i | \mathbf{y}^{(i)})}}$$

ここで、 $\log \widehat{DOR}_i$ は研究 i についての診断オッズ比の観測値を対数変換した値である。また、 $E_{\log DOR_i | \mathbf{y}^{(i)}}(\log DOR_i | \mathbf{y}^{(i)})$ および $\text{Var}_{\log DOR_i | \mathbf{y}^{(i)}}(\log DOR_i | \mathbf{y}^{(i)})$ は、それぞれ、診断オッズ比の事後予測平均および事後予測分散であり、MCMCより求めることができる。

なお、感度や偽陽性率の計算の場合と同様であるが、leave-one-out データセットからは研究 i が除かれているため、 $\log DOR_i$ の研究内分散についてはMCMCによって推定することができない。そのため、 $\log DOR_i$ の研究内分散は以下の式で推定する[22]。

$$\frac{1}{TP_i} + \frac{1}{TN_i} + \frac{1}{FP_i} + \frac{1}{FN_i}$$

4章についての解析プログラムは、付録Cにまとめる。

第5章 ベイズ流 P 値

5.1 背景と目的

3つ目の提案手法は、ベイズ流 P 値である。ベイズ流 P 値は事後予測分布を用いたモデル確認の方法であり[14, 23]、観測データと、想定されるモデルの間の乖離を評価する方法である。ベイズ流 P 値による外れ値の検出[14]は以下の流れで行われる。

1. ベイズ流 P 値を定義するにあたり、観測値 y_i と、想定されるモデルの間の乖離の指標 $D(y_i, \theta)$ を定義する。ここで、 θ はパラメータ全体を表す。

$$D(y_i, \theta) = \frac{[y_i - E_{y_i|\theta}(y_i|\theta)]^2}{\text{Var}_{y_i|\theta}(y_i|\theta)}$$

なお、乖離の指標 $D(y_i, \theta)$ については、上記以外の定義も可能である。

2. 事後予測サンプル y_i^* に対しても、乖離の指標 $D(y_i^*, \theta)$ を求める。
3. 事後予測サンプルに対する、観測データの極端さの指標として、ベイズ流 P 値を定義する。

$$P_D = P[D(y_i^*, \theta) > D(y_i, \theta) | \mathbf{y}]$$

観測値 y_i が極端な値であるほど、観測値に対する乖離の指標 $D(y_i, \theta)$ は大きくなり、その結果として、ベイズ流 P 値は小さくなる。このため、ベイズ流 P 値を用いて外れ値を検出することが可能である。

本章では、最初に、ベイズ流 P 値を診断法のメタアナリシスに適用することにより、感度または偽陽性率のどちらか一方のみを評価するベイズ流 P 値を提案する。次に、感度と偽陽性率を同時に評価するベイズ流 P 値として、Zhang et al. [15]によって提案された、乖離の指標の平均値を用いる指標を、診断法のメタアナリシスに適用する。さらに、感度および偽陽性率を統合し、その相関を考慮

したベイズ流 P 値を提案する。最後に、診断法のメタアナリシスのための要約指標を用いる方法として、診断オッズ比を用いるベイズ流 P 値を提案する。

5.2 感度または偽陽性率に対するベイズ流 P 値

最初に、研究 i の感度および偽陽性率についての周辺的なベイズ流 P 値を定義する。ベイズ流 P 値を導入するにあたり、感度および偽陽性率に対する乖離の指標として、それぞれ次の指標を定義する。

$$D_i^A(y_{Ai}, \xi) = \frac{[y_{Ai} - E_{y_{Ai}|\xi}(y_{Ai}|\xi)]^2}{\text{Var}_{y_{Ai}|\xi}(y_{Ai}|\xi)}, D_i^B(y_{Bi}, \xi) = \frac{[y_{Bi} - E_{y_{Bi}|\xi}(y_{Bi}|\xi)]^2}{\text{Var}_{y_{Bi}|\xi}(y_{Bi}|\xi)}$$

ここで $E_{y_{Ai}|\xi}(y_{Ai}|\xi)$ および $E_{y_{Bi}|\xi}(y_{Bi}|\xi)$ は事後予測平均であり、 $\text{Var}_{y_{Ai}|\xi}(y_{Ai}|\xi)$ および $\text{Var}_{y_{Bi}|\xi}(y_{Bi}|\xi)$ は事後予測分散である。また、 ξ はパラメータ全体を示すベクトル $\xi = (\boldsymbol{\mu}^T, \sigma_A^2, \sigma_B^2, \rho)^T$ である。次に、仮想的な観測値 y_{Ai}^* および y_{Bi}^* を用いて、乖離の指標 $D_i^A(y_{Ai}^*, \xi)$ および $D_i^B(y_{Bi}^*, \xi)$ を同様に定義する。仮想的な観測値 y_{Ai}^* および y_{Bi}^* は事後予測サンプルであり、MCMC によって生成される。ベイズ流 P 値は観測データの極端さの指標であり、研究 i の感度および偽陽性率についての周辺的なベイズ流 P 値は、上記の乖離の指標を用いることにより、それぞれ以下の式で定義される。

$$P_{D_i^A} = P[D_i^A(y_{Ai}^*, \xi) > D_i^A(y_{Ai}, \xi) | \mathbf{y}] = \int P[D_i^A(y_{Ai}^*, \xi) > D_i^A(y_{Ai}, \xi) | \xi] p(\xi | \mathbf{y}) d\xi,$$

$$P_{D_i^B} = P[D_i^B(y_{Bi}^*, \xi) > D_i^B(y_{Bi}, \xi) | \mathbf{y}] = \int P[D_i^B(y_{Bi}^*, \xi) > D_i^B(y_{Bi}, \xi) | \xi] p(\xi | \mathbf{y}) d\xi$$

このベイズ流 P 値が小さい場合、例えば < 0.15 や < 0.10 となる場合、観測値は事後予測サンプルと比較して、事後予測平均からの乖離が大きい、極端な値であったことが示される。そのため、小さいベイズ流 P 値に該当する研究は外れ

値であると考えられる。

5.3 感度と偽陽性率を同時に評価するベイズ流 P 値

本節では、感度および偽陽性率を同時に評価する指標について提案する。まず、周縁的な乖離の指標の平均値を用いてベイズ流 P 値を定義する方法が Zhang et al. [15]によって提案されており、この方法を診断法のメタアナリシスに適用することができる。

$$AD_i(\mathbf{y}_i, \xi) = \frac{1}{2} \left(\frac{[y_{Ai} - E_{y_{Ai}|\xi}(y_{Ai}|\xi)]^2}{\text{Var}_{y_{Ai}|\xi}(y_{Ai}|\xi)} + \frac{[y_{Bi} - E_{y_{Bi}|\xi}(y_{Bi}|\xi)]^2}{\text{Var}_{y_{Bi}|\xi}(y_{Bi}|\xi)} \right),$$

$$P_{AD_i} = P[AD_i(\mathbf{y}_i^*, \xi) > AD_i(\mathbf{y}_i, \xi) | \mathbf{y}] = \int P[AD_i(\mathbf{y}_i^*, \xi) > AD_i(\mathbf{y}_i, \xi) | \xi] p(\xi | \mathbf{y}) d\xi$$

上記の、乖離の指標を平均する方法は、感度と偽陽性率の両面で、事後予測平均からの乖離を測る指標となっているものの、感度と偽陽性率の間の相関が考慮されていない点で不十分な指標であると考えられる。

従って、本研究では、相対距離や標準化残差での定義と同様に、感度および偽陽性率を統合し、その相関を考慮したベイズ流 P 値を提案する。そのために、乖離に関する 2 次元の指標を定義する。

$$SD_i(\mathbf{y}_i, \xi) = (\mathbf{y}_i - E_{\mathbf{y}_i|\xi}(\mathbf{y}_i|\xi))^T V_{\mathbf{y}_i|\xi}(\mathbf{y}_i|\xi)^{-1} (\mathbf{y}_i - E_{\mathbf{y}_i|\xi}(\mathbf{y}_i|\xi))$$

そして、上記の 2 次元の乖離の指標を用いて、ベイズ流 P 値を定義する。

$$P_{SD_i} = P[SD_i(\mathbf{y}_i^*, \xi) > SD_i(\mathbf{y}_i, \xi) | \mathbf{y}] = \int P[SD_i(\mathbf{y}_i^*, \xi) > SD_i(\mathbf{y}_i, \xi) | \xi] p(\xi | \mathbf{y}) d\xi$$

この、統合されたベイズ流 P 値は、2 次元空間での乖離の情報を反映させる指標であり、診断法のメタアナリシスの影響力の評価に適した指標である。

5.4 診断オッズ比を用いたベイズ流 P 値

ベイズ流 P 値に関する最後の提案法として、診断オッズ比を用いたベイズ流 P 値を提案する。以下に、ベイズ流 P 値と乖離の指標を示す。

$$P_{D_{DORi}} = P[D(\log \widehat{DOR}_i^*, \xi) > D(\log \widehat{DOR}_i, \xi) | \mathbf{y}],$$

$$D_{DORi}(\log \widehat{DOR}_i, \xi) = \frac{[\log \widehat{DOR}_i - E_{\log DOR_i | \xi}(\log DOR_i | \xi)]^2}{Var_{\log DOR_i | \xi}(\log DOR_i | \xi)}$$

$\log \widehat{DOR}_i$ は、研究 i の診断オッズ比（観測値）の対数変換値であり、 $\log \widehat{DOR}_i^*$ は、研究 i の診断オッズ比（事後予測サンプル）の対数変換値である。研究 i についての $\log(DOR)$ の事後予測平均と事後予測分散は MCMC から得られる。

5 章についての解析プログラムは、付録 D にまとめる。

第6章 SROC 曲線の AUC への影響度

6.1 背景と目的

4つ目の提案手法は、SROC 曲線の曲線下面積 (AUC) を用いる影響度診断である。診断法のメタアナリシスにおいて、SROC 曲線の AUC は診断性能に関する要約指標として、広く用いられている。具体的には AUC の値が 1.0 に近いほど診断性能は良く、0.5 に近いほど診断性能は悪いとされている。そのため、SROC 曲線の AUC は、それぞれの研究がどの程度の影響力をもつかを評価するための有用なツールとなり、AUC を用いた影響度診断を行うことができると考える。

本章では、 i 番目の研究をメタアナリシスから除外する leave-one-out 交差検証法によって、SROC 曲線の AUC の推定値への影響を評価する指標を提案する。

6.2 SROC 曲線の AUC への影響度

本節では、SROC 曲線の AUC への影響度として、以下の指標を提案する。

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

ここで、 \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。 ΔAUC_i の絶対値が大きくなる研究は、SROC 曲線の形状および AUC の観点から、影響力が大きいと考えられる。

ΔAUC_i を求める際には、SROC 曲線を推定し、その AUC を求める必要がある。本研究においては、2 章の 2 変量メタアナリシスでの推定値を用いることにより、SROC 曲線および AUC を次のように推定する[4]。

1. ロジット変換後の感度および偽陽性率についての回帰直線を以下の式により推定する。

$$E[Y_A|Y_B] = \hat{\mu}_A + \frac{\hat{\rho}\hat{\sigma}_A\hat{\sigma}_B}{\hat{\sigma}_B^2}(Y_B - \hat{\mu}_B)$$

2. 回帰直線をロジット変換後のスケールから、0~1 のスケールに逆ロジット変換することにより SROC 曲線を推定する。
3. SROC の AUC を台形法によって推定する[2, 5]。

上記の SROC 曲線の推定方法は、2 変量メタアナリシスモデルに関する代表的な文献である Reitsma et al. [4]に従った。しかしながら、上記以外の SROC 曲線の推定方法に対しても、提案法を用いることは可能である。

AUC の計算においては 0 から 1 の範囲で AUC を推定する方法と、データが存在する偽陽性率の範囲で計算した部分的な AUC の値を求めてから、その部分的な AUC を補正することにより、0 から 1 の範囲での AUC を推定する方法がある。後者の方法については、leave-one-out 交差検証法に伴い、部分的な AUC を求める際の偽陽性率の範囲が変わることがあり、それに伴い AUC の値が変化するものの、この変化を検出することは本論文の目的に沿わない。従って、本論文では前者の 0 から 1 の範囲で AUC を推定する方法を採用する。

なお、AUC と診断オッズ比の間には正の相関が存在していると一般的に考えられるため、AUC で評価する影響力と診断オッズ比で評価する影響力の間には正の相関が存在すると考えられる。

6 章についての解析プログラムは、付録 B および C にまとめる。

第7章 尺度混合正規分布における尺度パラメ

一夕の事後推定値

7.1 背景と目的

5つ目の提案手法は、尺度混合正規分布における尺度パラメータの事後推定値である。尺度混合正規分布は Andrews and Mallows [24]によって提案された手法であり、Carlin and Louis [14]によってベイズモデルにおける外れ値の検出に用いられている。尺度混合正規分布を用いることによって、正規分布に従う誤差を、Student の t 分布に従う誤差のような、裾の重い誤差に拡張することができる。

最初に、尺度混合正規分布について説明する。例示として、以下のモデルを考える。

$$y_i = \mu + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2), \quad i = 1, \dots, n$$

Andrews and Mallows [24]によって、誤差の分布に、尺度パラメータ λ_i を導入することが提案された。上記のモデルに対して、尺度パラメータ λ_i を導入したモデルを以下に示す。

$$y_i = \mu + \epsilon_i, \quad \epsilon_i \sim N(0, \lambda_i \sigma^2), \quad i = 1, \dots, n$$

そして、尺度パラメータに事前分布を設定することで、誤差の分布を、t分布のような、裾が重く、外れ値を許容する分布に拡張することが可能となる。尺度パラメータの事前分布と、誤差の分布としては、表 7.1 の組み合わせが知られている [14]。

表 7.1 尺度パラメータの事前分布と誤差の分布

尺度パラメータの事前分布	誤差の分布
$\lambda_i = 1$	正規分布
$\lambda_i \sim IG(v/2, 2/v)$	t分布 (自由度 v)
$\lambda_i \sim EXP(2)$	二重指数分布

IG：逆ガンマ分布，EXP：指数分布

実践的なデータにおいては、正規分布よりも裾の重い誤差分布が認められる場合があり、そのような場合に尺度混合正規分布は有用である。

一方で、極端な観測値に対しては、尺度パラメータ λ_i の推定値が大きくなるため、尺度パラメータの事後推定値を用いて、外れ値の検出を行うことができる [14]。Zhang et al. [15]は尺度混合正規分布をネットワークメタアナリシスへ適用し、尺度混合正規分布の尺度パラメータ λ_i の事後推定値が 1 を上回る確率を求め、その確率が大きくなる場合を外れ値とした。

本章では、尺度混合正規分布による外れ値の検出について、診断法のメタアナリシスへの適用を提案する。また、感度と偽陽性率について、1つのパラメータで外れ値を評価する指標を提案する。

7.2 感度または偽陽性率に対する尺度パラメータ

本節では、Carlin and Louis [14]で採用されたアプローチを、診断法のメタアナリシスに適用し、感度および偽陽性率のそれぞれに対して外れ値を検出するための指標を提案する。

Carlin and Louis [14]の方法を適用するために、本節では、2章で提示したモデルとは異なるモデルを用いることとする。本節で用いるモデルは、診断法のメタ

アナリシスでしばしば用いられている Reitsma et al. [4]により提案されたモデル（ロジット変換された感度および偽陽性率の推定量の漸近的な正規近似に基づくモデル）の拡張である。

まず、以下の2項確率モデルを考える。

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

ここで、 p_{Ai} および p_{Bi} は、それぞれ研究 i の感度および偽陽性率である。そして、ロジット変換され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 $y_{Bi} = \text{logit}(FP_i/n_{Bi})$ とおき、 y_{Ai} および y_{Bi} についての2変量正規分布を考える。

$$\begin{pmatrix} y_{Ai} \\ y_{Bi} \end{pmatrix} \sim N \left(\begin{pmatrix} \theta_{Ai} \\ \theta_{Bi} \end{pmatrix}, \begin{pmatrix} \lambda_{Ai}\omega_{Ai}^2 & 0 \\ 0 & \lambda_{Bi}\omega_{Bi}^2 \end{pmatrix} \right)$$

θ_{Ai} および θ_{Bi} は、研究 i についての、ロジット変換された感度および偽陽性率の平均パラメータである。 λ_{Ai} および λ_{Bi} は尺度混合正規分布を特徴づける尺度パラメータである。また、 ω_{Ai}^2 および ω_{Bi}^2 は研究内分散パラメータであり、4.2節と同様に、以下の式で推定値が得られる[4]。

$$\omega_{iA}^2 = \frac{1}{(TP_i + FN_i) \cdot \frac{TP_i}{(TP_i + FN_i)} \cdot \frac{FN_i}{(TP_i + FN_i)}}$$

$$\omega_{iB}^2 = \frac{1}{(TN_i + FP_i) \cdot \frac{TN_i}{(TN_i + FP_i)} \cdot \frac{FP_i}{(TN_i + FP_i)}}$$

次に、ロジット変換された2項確率パラメータについての変量効果モデルを考える。

$$\begin{pmatrix} \theta_{Ai} \\ \theta_{Bi} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, \begin{pmatrix} \sigma_A^2 & \rho\sigma_A\sigma_B \\ \rho\sigma_A\sigma_B & \sigma_B^2 \end{pmatrix} \right)$$

ここで μ_A および μ_B は全体の平均パラメータであり、 σ_A 、 σ_B 、および ρ は異質性の分散パラメータおよびその相関パラメータである。これらのパラメータ

について適切な事前分布を与えることにより、各パラメータの事後推定値をMCMCによって得ることができる。本論文においては、尺度パラメータ λ_{Ai} および λ_{Bi} の事前分布として、表 7.1 で示した逆ガンマ分布 $\lambda_{Ai}, \lambda_{Bi} \sim IG(\nu/2, 2/\nu)$ および指数分布 $\lambda_{Ai}, \lambda_{Bi} \sim EXP(2)$ の、2種類の分布を用いることとした。

そして、尺度パラメータの事後推定値を用いて、正規誤差の仮定の妥当性を評価する。具体的には、 λ_{Ai} または λ_{Bi} の事後推定値が 1 に等しい場合の誤差分布は正規分布と考えられるが、 λ_{Ai} または λ_{Bi} が 1 より大きい場合の誤差分布は正規分布よりも裾の重い分布と考えられる。従って、ある研究について、尺度パラメータ λ_{Ai} または λ_{Bi} の事後推定値が 1 よりも大きくなる確率を求め、その確率が大きい研究を外れ値として検出する。

7.3 感度と偽陽性率を同時に評価する尺度パラメータ

本節では、感度および偽陽性率を同時に評価する尺度パラメータについて提案する。

7.2 節の方法は、感度および偽陽性率について別々の尺度パラメータを設定し、感度および偽陽性率のそれぞれに対して外れ値を検出する指標であった。本節では、7.2 節の尺度パラメータに $\lambda_i = \lambda_{Ai} = \lambda_{Bi}$ を仮定し、共通の尺度パラメータ (common scale parameter, CSP) λ_i の事後推定を行う方法を提案し、この方法を CSP モデルと呼ぶこととする。CSP モデルにおいては、研究内および研究間のモデルは、以下のように書き表せる。

$$\begin{pmatrix} y_{Ai} \\ y_{Bi} \end{pmatrix} \sim N \left(\begin{pmatrix} \theta_{Ai} \\ \theta_{Bi} \end{pmatrix}, \begin{pmatrix} \lambda_i \omega_{Ai}^2 & 0 \\ 0 & \lambda_i \omega_{Bi}^2 \end{pmatrix} \right)$$

$$\begin{pmatrix} \theta_{Ai} \\ \theta_{Bi} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_A \\ \mu_B \end{pmatrix}, \begin{pmatrix} \sigma_A^2 & \rho \sigma_A \sigma_B \\ \rho \sigma_A \sigma_B & \sigma_B^2 \end{pmatrix} \right)$$

このモデルを用いることにより、結果として得られる λ_i の事後推定値は、感度および偽陽性率の両方の点から、研究 i についての正規誤差からの乖離を反映することとなる。CSP モデルを用いることにより、感度と偽陽性率に対する 2 つのパラメータでなく、1 つのパラメータによって外れ値を評価することができるという利点が得られる。

7 章についての解析プログラムは、付録 E にまとめる。

第8章 実データへの適用（小児の尿路感染症

患者における膀胱尿管逆流の超音波によるスクリーニング)

8.1 適用データの紹介

本章では、3章から7章で提案された指標の例示として、小児の尿路感染症患者における膀胱尿管逆流の超音波によるスクリーニングのデータに提案法を適用する。この診断法のメタアナリシスは Shaikh et al. [17] によって実施されたものである。膀胱尿管逆流（特に重症の場合）は、尿路感染症および腎瘢痕化のリスク上昇につながるため、膀胱尿管逆流の検出についての関心は高い。しかしながら、標準的な検査法とされている排泄時膀胱尿道造影法は、膀胱カテーテルを必要とし、小児への放射線曝露を伴うため、スクリーニングのための画像診断の代替法が研究されている。Shaikh et al. [17] は、2つの代替法（ジメルカプトコハク酸による造影、および、超音波）について、診断法のメタアナリシスを実施した。

本論文では、2つのメタアナリシスのうち、外れ値が明らかに認められると考えられる、超音波によるスクリーニングのデータに提案法を適用する。この診断法のメタアナリシスには20研究が含まれており、それらの個別研究データを表8.1に示す。また、感度および偽陽性率の散布図およびSROC曲線を図8.1に示す。図8.1より、研究間の異質性が大きいことが読み取れるが、異質性のいくらかの部分については、小児用の採尿バッグの使用、および、発熱の認められない小児の研究への参加によって説明することができる[17]。

本事例においては、図 8.1 より、研究 1 が外れ値であると考えられる。研究 1 は最も古い 1986 年に公表されており、最も少ない参加者数 (N=29) で実施された。研究が実施された時期と、少ない参加者数に伴う大きな偶然変動によって、研究 1 が外れ値になった可能性がある。また、研究 9 も、偽陽性率が低いにも関わらず、感度が高いので、外れ値であると考えられる。研究 15 については、85% という極端に高い偽陽性率をもつ点に関して、極端なプロファイルをもつ研究と考えられる。本事例への適用においては、研究 1, 9, 15 の検出についての性能確認を行い、また、それ以外にも、結果に強い影響を与える研究の検出についての性能評価を行う。

表 8.1 小児の尿路感染症患者における膀胱尿管逆流の超音波によるスクリーニングのデータ

番号	研究	TP	FP	FN	TN	感度	偽陽性率
						[95%信頼区間]	[95%信頼区間]
1	Alon (1986)	7	9	11	2	0.39 [0.17, 0.64]	0.82 [0.48, 0.98]
2	Alon (1999)	8	2	30	60	0.21 [0.10, 0.37]	0.03 [0.00, 0.11]
3	Calisti (2005)	26	31	19	71	0.58 [0.42, 0.72]	0.30 [0.22, 0.40]
4	Cleper (2004)	6	7	7	44	0.46 [0.19, 0.75]	0.14 [0.06, 0.26]
5	El Shenoufy (2009)	5	2	8	30	0.38 [0.14, 0.68]	0.06 [0.01, 0.21]
6	Goldman (2000)	8	4	13	19	0.38 [0.18, 0.62]	0.17 [0.05, 0.39]
7	Hoberman (2003)	14	21	101	165	0.12 [0.07, 0.20]	0.11 [0.07, 0.17]
8	Ismaili (2011)	18	21	35	135	0.34 [0.22, 0.48]	0.13 [0.09, 0.20]
9	Kim (2006)	9	0	6	37	0.60 [0.32, 0.84]	0.00 [0.00, 0.09]
10	Lee (2009)	47	44	20	109	0.70 [0.58, 0.81]	0.29 [0.22, 0.37]
11	Lee (2012)	72	141	49	356	0.60 [0.50, 0.68]	0.28 [0.24, 0.33]
12	Lopez Sastre (2007)	25	55	29	118	0.46 [0.33, 0.60]	0.32 [0.25, 0.39]
13	Mahant (2002)	14	30	21	97	0.40 [0.24, 0.58]	0.24 [0.17, 0.32]
14	Montini (2009)	18	20	48	214	0.27 [0.17, 0.40]	0.09 [0.05, 0.13]
15	Morin (1999)	20	41	2	7	0.91 [0.71, 0.99]	0.85 [0.72, 0.94]
16	Muller (2009)	29	91	23	147	0.56 [0.41, 0.70]	0.38 [0.32, 0.45]
17	Sheu (2013)	60	70	93	250	0.39 [0.31, 0.47]	0.22 [0.17, 0.27]
18	Soylu (2007)	5	4	33	46	0.13 [0.04, 0.28]	0.08 [0.02, 0.19]
19	Supavekin (2013)	9	14	8	36	0.53 [0.28, 0.77]	0.28 [0.16, 0.42]
20	Tsai (2012)	40	96	20	64	0.67 [0.53, 0.78]	0.60 [0.52, 0.68]

TP : true positive 真陽性, FP : false positive 偽陽性, FN : false negative 偽陰性,
TN : true negative 真陰性

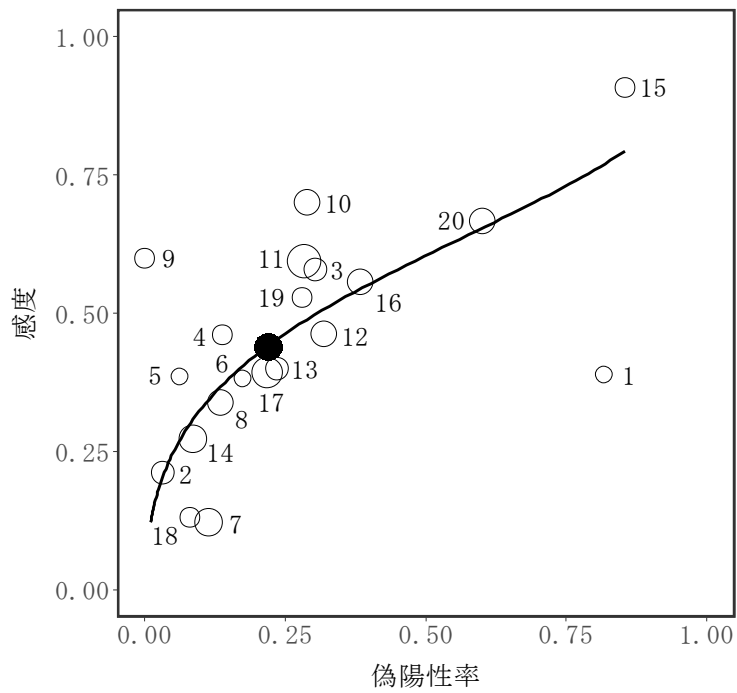


図 8.1 小児の尿路感染症患者における膀胱尿管逆流の超音波によるスクリーニングのデータ（散布図および SROC 曲線）

（○：個別研究の感度および偽陽性率 [印の大きさは参加者数を反映する]，

●：診断法のメタアナリシスにより推定された全体の感度および偽陽性率）

8.2 計算プログラムと実行条件

実データへの提案法の適用においては、OpenBUGS version 3.2.3 [20, 21]および R version 3.5.1 [25]を用いて、プログラムを実装した。また、MCMC については 120,000 回の反復計算を実施し、そのうちの最初の 20,000 回については burn-in とした。

反復計算の回数については、実行結果が十分に安定し、計算時間が許容できるという観点で設定しており、この回数が最適というものではない。なお、診断法のメタアナリシスへの適用ではないが、Carlin and Louis [14]および Zhang et al. [15]では、反復計算回数は 1,000 回 (burn-in を除く) であった。しかしながら、今回の適用で反復計算回数を 1,000 回とすると、外れ値を検出するための閾値に近い研究に対して、プログラムを実行するたびに、外れ値に該当するか否かが変わる結果となった。そのため、実行結果を安定させる観点から、余裕をもたせた反復計算回数とした。なお、MCMC の収束に関しては、パラメータの推定値の推移を示すプロットにより確認し、問題は認められなかった。

leave-one-out 交差検証法に基づく各指標の計算においては、leave-one-out データセットを R で作成して、OpenBUGS で MCMC の計算を行うというプロセスを繰り返すことにより実行した。つまり、8.1 節のデータ (20 研究) に対しては、OpenBUGS が 20 回実行されている。なお、Carlin and Louis [14]においては、OpenBUGS の実行が 1 回で済むようなプログラムコードも記載されており、本研究においては計算時間の問題が生じなかったものの、反復計算の回数や、プログラムコードについては改良の余地がある。

計算には市販のノート PC を用いており (OS : Windows 10, CPU : Intel® Core™ i5, メモリ : 8 GB), 全ての提案法に関するプログラムを実行して、計算時間は合計で 1 時間程度であった。なお、計算時間を節約するために、leave-one-out

交差検証法を用いる指標（相対距離，標準化残差，SROC 曲線の AUC への影響度）については，leave-one-out データセットの作成と MCMC によるパラメータの推定はまとめて行い，その後，得られた推定値を用いて，各指標の算出を行うこととした（付録 C を参照）。

なお，本節の内容に関しては，9 章の実データへの適用（喘息患者における好酸球性気道炎症の代替マーカー）についても同様である。そのため，9 章では，計算プログラムと実行条件に関する記載は省略する。

8.3 提案法の適用結果

相対距離

図 8.2 に相対距離の結果を示す。図 8.2 の縦軸は相対距離の値であり，横軸は leave-one-out 交差検証法によって除外された研究の番号である。図 8.2 には 5 個のパネルがあり，(a) 感度についての相対距離，(b) 偽陽性率についての相対距離，(c) 統合相対距離，(d) 平均相対距離，(e) 診断オッズ比についての相対距離を示す。

まず，感度と偽陽性率のそれぞれに対する相対距離の結果を述べる。相対距離（感度）については，高い感度をもつ研究 15，低い感度をもつ研究 7 について，0.05 を超える絶対値となった。相対距離（偽陽性率）については，高い偽陽性率をもつ研究 15 に対して，0.10 を超える絶対値となった。なお，同程度の感度や偽陽性率であっても，参加者数が多い研究の方が，全体の推定値へ与える影響が大きかった。具体的には，研究 1 と 15 は同程度の偽陽性率をもつが，参加者数が多い研究 15 の方が相対距離（偽陽性率）は大きな値となった。また，相対距離（感度）と相対距離（偽陽性率）を比べると，全体的に，相対距離（偽陽性率）の方が，絶対値が大きくなっている。これは，相対距離を求める際の分母（感度

または偽陽性率の全体の推定値)が異なり、偽陽性率の方が、分母が小さくなるためである。

次に、統合相対距離に関して述べると、研究 7, 15 のみが 0.05 よりも大きな統合相対距離をもち、これらの研究は影響力の大きな研究であることが示唆された。研究 15 は感度および偽陽性率の両方について最大の値をもつ研究であり、研究 7 は最も低い感度をもちつつ参加者数が多い研究である。一方で、平均相対距離に関しては、研究 2, 15, 18 について、0.05 よりも大きな平均相対距離となった。研究 2 と 18 は、感度と偽陽性率の両方に関して、低い値をもつ研究である。統合相対距離と平均相対距離については、1次元のユークリッド距離を平均した指標か、2次元空間でのユークリッド距離を用いた指標かという違いがあり、検出される研究は閾値に依存するものの、研究 2, 7, 15, 18, 20 に対して、大きな値となる点で共通していた。

最後に、診断オッズ比を用いた相対距離の結果を述べる。診断オッズ比に関しては、図 8.2 より、研究 1, 7, 9, 10 が 0.05 より大きな絶対値となった。診断オッズ比を用いた結果については、感度が高く偽陽性率が低い場合に診断オッズ比は大きくなり、感度が低く偽陽性率が高い場合に診断オッズ比は小さくなるという、診断オッズ比の性質のため、図 8.1 で、左上または右下方向に外れている研究が検出された。

相対距離の適用結果としては、感度または偽陽性率への影響度が大きい研究として研究 2, 7, 15, 18, 診断オッズ比への影響度が大きい研究として研究 1, 7, 9, 10 が認められた。

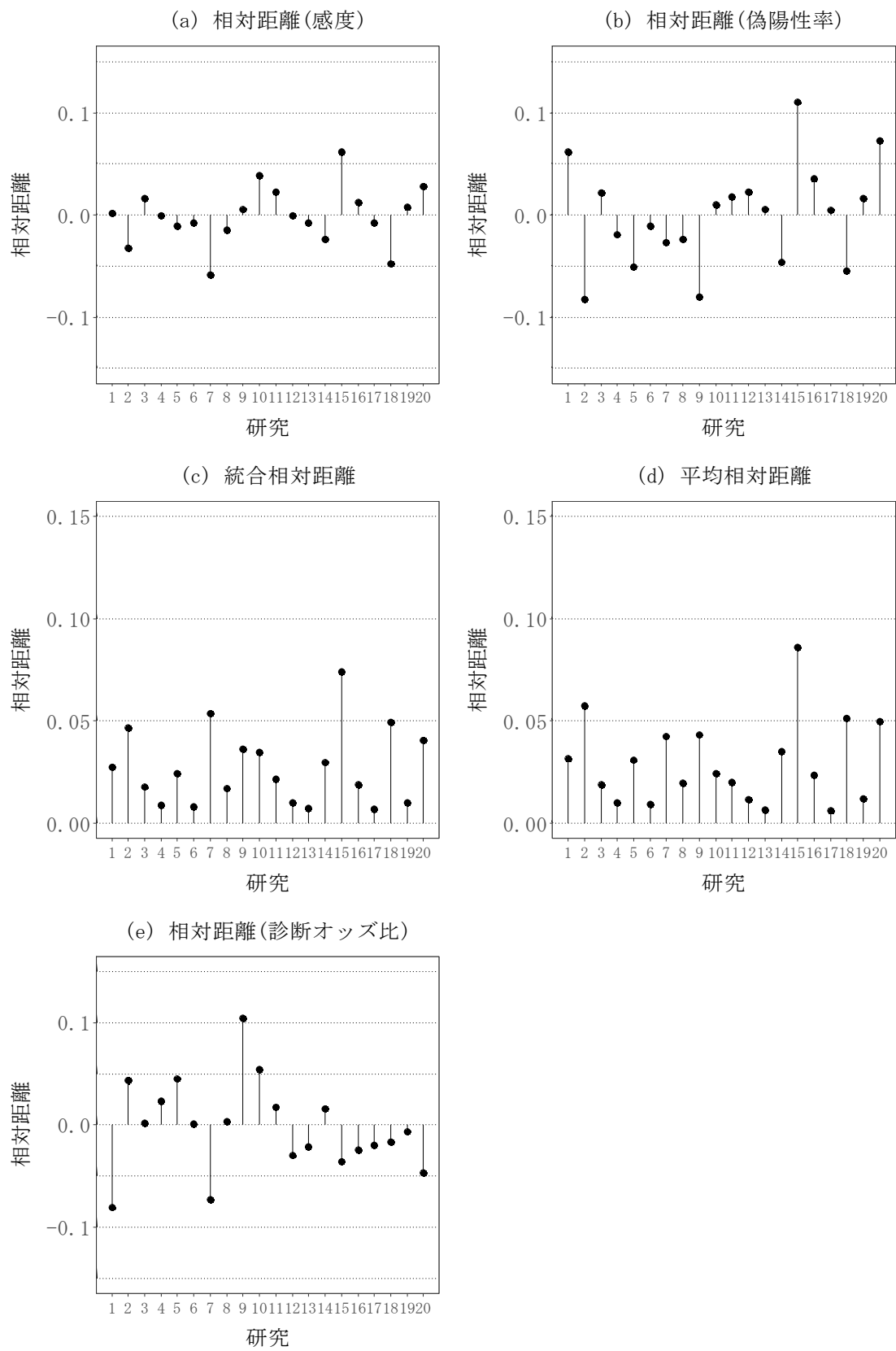


図 8.2 相対距離の結果

標準化残差

図 8.3 に標準化残差の結果を示す。図 8.3 の縦軸は標準化残差の値であり、横軸は leave-one-out 交差検証法によって除外された研究の番号である。図 8.3 には 5 個のパネルがあり、(a) 感度についての標準化残差、(b) 偽陽性率についての標準化残差、(c) 統合標準化残差、(d) 平均標準化残差、(e) 診断オッズ比についての標準化残差を示す。

まず、感度と偽陽性率のそれぞれに対する標準化残差の結果を述べる。標準化残差（感度）については、高い感度をもつ研究 15、低い感度をもつ研究 7 と 18 について、 ± 1.64 の範囲外の値となり、これらの研究が外れ値であることが示唆された。 ± 1.64 という閾値は、標準正規分布の上側 5% 点 / 下側 5% 点である。また、標準化残差（偽陽性率）については、高い偽陽性率をもつ研究 1、15 に対して、また、低い偽陽性率をもつ研究 9 について、 ± 1.64 の範囲外の値となった。

次に、統合標準化残差に関して述べると、研究 1、7、9、15 が 4.61 より大きな統合標準化残差をもち、これらの研究が外れ値であることが示唆された。4.61 という閾値は、自由度 2 のカイ 2 乗分布の上側 10% 点である。研究 1、9、15 は大きな、または、小さな偽陽性率をもつ研究であり、研究 7、15 は大きな、または、小さな感度をもつ研究である。平均標準化残差に関しては、研究 15 について大きな平均標準化残差となったものの、外れ値と考えられる研究 1 と 9 に対しては、平均標準化残差は大きな値とならなかった。逆に、感度と偽陽性率の両面について中程度に外れている研究 2 や 18 の方が、大きな平均標準化残差となった。標準化残差については、外れ値と考えられる研究 1 と 9 の検出結果の点で、平均標準化残差よりも統合標準化残差の方が良い性能を示した。

最後に、診断オッズ比を用いた標準化残差の結果を述べる。診断オッズ比に関しては、図 8.3 より、研究 1、9 の標準化残差が ± 1.64 の範囲外の値となった。相対距離の場合と同様であるが、診断オッズ比の性質のため、図 8.1 で、左上また

は右下方向に外れている研究が検出された。

標準化残差の適用結果としては、外れ値の可能性のある研究として、研究 1, 7, 9, 15, 18 が認められた。

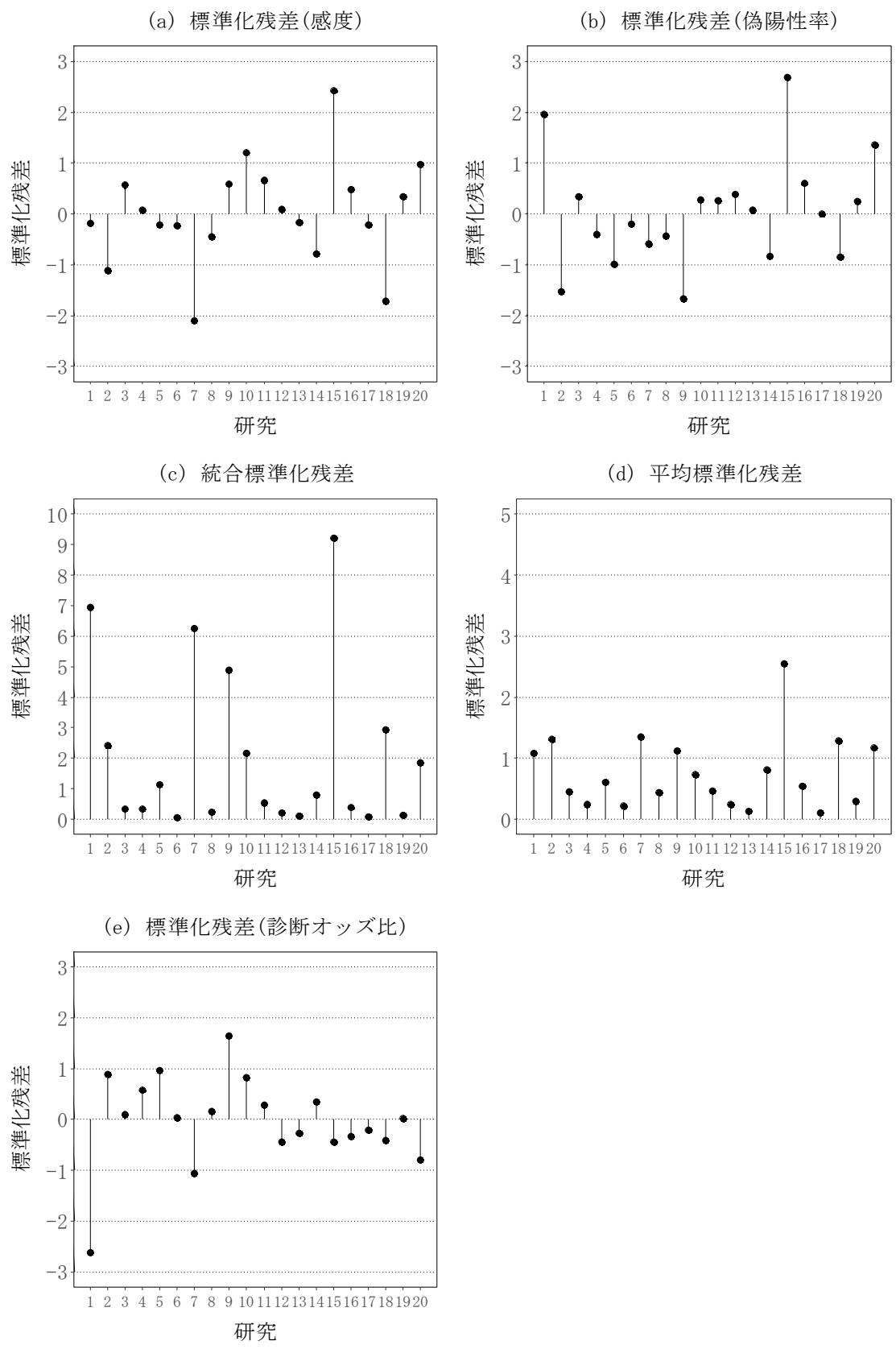


図 8.3 標準化残差の結果

ベイズ流 P 値

表 8.2 にベイズ流 P 値（感度，偽陽性率，感度および偽陽性率を統合した指標，感度および偽陽性率の平均に相当する指標，診断オッズ比）を示す。研究 1, 9 は，偽陽性率，統合した指標，平均に相当する指標および診断オッズ比について，小さなベイズ流 P 値 (<0.15) となり，その中のいくつかについては 0.05 よりも小さくなった。それゆえに，研究 1, 9 はベイズ流 P 値の点からは，外れ値の可能性があると判定された。図 8.1 より，研究 1, 9 は SROC 曲線で示される感度および偽陽性率の傾向から外れた研究であることが分かる。また，研究 1, 9 を除くと，0.3 よりも小さなベイズ流 P 値は全く認められなかった。

感度および偽陽性率を統合した指標と，感度および偽陽性率の平均に相当する指標の比較を行う。外れ値と考えられる研究 1, 9 に対しては，統合した指標の方が，平均に相当する指標よりも，ベイズ流 P 値が小さくなった。今回の適用事例に対しては，どちらの指標を用いても研究 1 と 9 は外れ値と検出される結果となったが，統合した指標の方が小さいベイズ流 P 値となり，外れ値を強調する結果となった。

SROC 曲線の AUC への影響度

図 8.4 に，各研究をメタアナリシスから除外することによる，SROC 曲線の AUC への影響度解析の結果を示す。 ΔAUC_i は 20 研究のほとんどに対して，絶対値で 0.02 未満の値となったが，研究 1, 15 については，それぞれ， -0.036 ， 0.028 の値となった。これらの 2 つの研究は ΔAUC_i の点から，影響力が大きいことが示唆された。研究 15 は感度および偽陽性率の両方について最大の値をもつ研究であり，研究 1 は高い偽陽性率の割に低い感度をもつ研究であり，これらの研究は SROC 曲線の形状への影響が大きく，それに伴い AUC への影響度が大きくなったと考えられる。

表 8.2 ベイズ流 P 値の結果

番号	研究	感度	偽陽性率	統合指標	平均	診断 オッズ比
1	Alon (1986)	0.2916	0.0973	0.1263	0.1469	0.0415
2	Alon (1999)	0.9777	0.5548	0.8394	0.8388	0.5940
3	Calisti (2005)	0.8087	0.9840	0.9713	0.9713	0.8500
4	Cleper (2004)	0.6655	0.8979	0.8898	0.9017	0.6457
5	El Shenoufy (2009)	0.6640	0.5759	0.7194	0.7750	0.4150
6	Goldman (2000)	0.9668	0.9404	0.9957	0.9967	0.9231
7	Hoberman (2003)	0.3992	0.7125	0.6521	0.6575	0.3802
8	Ismaili (2011)	0.9707	0.9922	0.9992	0.9992	0.9727
9	Kim (2006)	0.1565	0.0228	0.0244	0.0288	0.0012
10	Lee (2009)	0.5452	0.8798	0.8188	0.8212	0.5566
11	Lee (2012)	0.8247	0.9683	0.9760	0.9761	0.8310
12	Lopez Sastre (2007)	0.8894	0.8950	0.9803	0.9814	0.8441
13	Mahant (2002)	0.8790	0.8990	0.9794	0.9808	0.8373
14	Montini (2009)	0.9641	0.9614	0.9980	0.9977	1.0000
15	Morin (1999)	0.3340	0.7052	0.6092	0.5808	0.5474
16	Muller (2009)	0.9914	0.9347	0.9968	0.9967	0.9770
17	Sheu (2013)	0.9459	0.9311	0.9939	0.9941	0.9126
18	Soylu (2007)	0.3935	0.7932	0.6298	0.6659	0.4081
19	Supavekin (2013)	0.8332	0.9390	0.9780	0.9760	0.8979
20	Tsai (2012)	0.9816	0.8696	0.9852	0.9854	0.9138

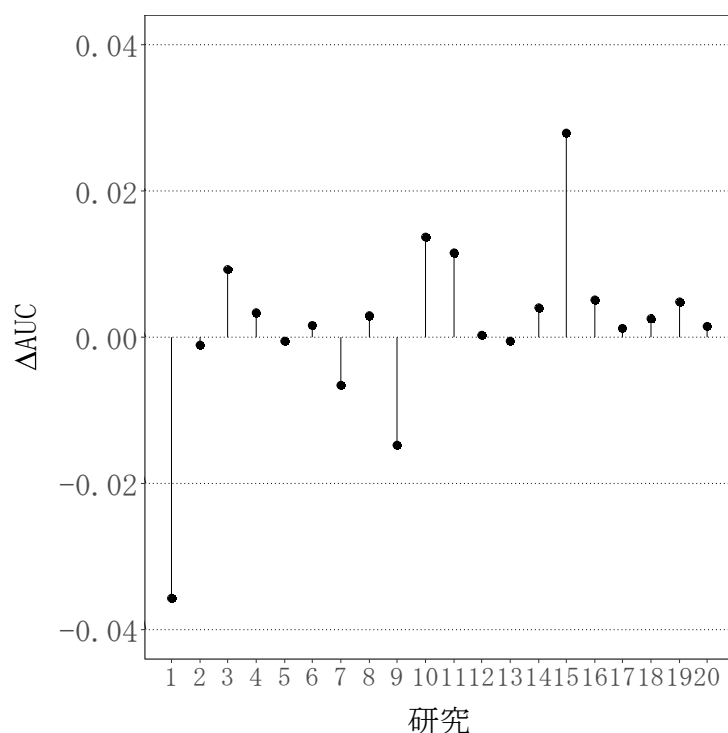


図 8.4 SROC 曲線の AUC への影響度の結果

尺度混合正規分布における尺度パラメータの事後推定値

表 8.3 に尺度混合正規分布における尺度パラメータの事後推定値に関する結果を示す。この表では、感度、偽陽性率および CSP モデルについて、尺度パラメータが 1 よりも大きくなる事後確率を示している。 λ_i の事前分布は Zhang et al. [15]と同じ条件を用いることとし、具体的には、逆ガンマ分布 $IG(1, 1)$ および指数分布 $EXP(2)$ を用いた。本適用では、尺度パラメータが 1 よりも大きくなる事後確率が 0.8 を上回る場合に外れ値とした。

適用結果として、まず、感度と偽陽性率のそれぞれに対する尺度パラメータの事後推定値について述べる。感度については、研究 7 について $\Pr(\lambda_i > 1) > 0.8$ となった。また、偽陽性率については、低い偽陽性率をもつ研究 9、高い偽陽性率をもつ研究 1 について $\Pr(\lambda_i > 1) > 0.8$ となった。

次に、CSP モデルの結果について述べる。CSP モデルについては、研究 1, 7,

9 について $\Pr(\lambda_i > 1) > 0.8$ となった。

この適用事例において、尺度パラメータ λ_i の事前分布の選択は、 $\Pr(\lambda_i > 1)$ の値にほとんど影響を及ぼさず、どちらの事前分布を用いても同じ研究が外れ値として検出された。

尺度混合正規分布における尺度パラメータの事後推定値の適用結果としては、外れ値の可能性のある研究として、研究 1, 7, 9 が認められた。

表 8.3 尺度混合正規分布における尺度パラメータの事後推定値（尺度パラメータが 1 よりも大きくなる事後確率）

番号	研究	感度		偽陽性率		CSP モデル	
		IG	EXP	IG	EXP	IG	EXP
1	Alon (1986)	0.6215	0.6313	0.9300	0.9241	0.9500	0.9560
2	Alon (1999)	0.5816	0.5655	0.6838	0.6706	0.6350	0.6268
3	Calisti (2005)	0.5668	0.5546	0.5929	0.5786	0.5492	0.5292
4	Cleper (2004)	0.5437	0.5310	0.6034	0.5974	0.5218	0.5365
5	El Shenoufy (2009)	0.5377	0.5310	0.6656	0.6642	0.5861	0.6192
6	Goldman (2000)	0.5315	0.5154	0.5315	0.5202	0.4414	0.4191
7	Hoberman (2003)	0.8575	0.8373	0.6217	0.6118	0.8685	0.8578
8	Ismaili (2011)	0.5541	0.5380	0.5890	0.5789	0.5166	0.5041
9	Kim (2006)	0.6680	0.7487	0.9742	0.9604	0.9936	0.9920
10	Lee (2009)	0.7756	0.7511	0.6414	0.6164	0.7991	0.7834
11	Lee (2012)	0.6532	0.6304	0.6239	0.6053	0.6591	0.6425
12	Lopez Sastre (2007)	0.5627	0.5581	0.6178	0.5997	0.5490	0.5547
13	Mahant (2002)	0.5413	0.5292	0.5943	0.5865	0.5193	0.5106
14	Montini (2009)	0.5661	0.5516	0.6165	0.5960	0.5534	0.5369
15	Morin (1999)	0.6875	0.6551	0.7687	0.7202	0.7917	0.7213
16	Muller (2009)	0.5432	0.5357	0.6145	0.6042	0.5345	0.5181
17	Sheu (2013)	0.5904	0.5749	0.6054	0.5926	0.5710	0.5659
18	Soylu (2007)	0.7150	0.7028	0.5660	0.5451	0.6786	0.6720
19	Supavekin (2013)	0.5113	0.4975	0.5686	0.5610	0.4725	0.4532
20	Tsai (2012)	0.5804	0.5660	0.6713	0.6411	0.6041	0.5934

IG : λ_i の事前分布として逆ガンマ分布を用いる場合, EXP : λ_i の事前分布として指数分布を用いる場合

8.4 メタアナリシスの結果に対する影響力の評価

最後に、全ての 20 研究を用いたメタアナリシスと、外れ値または影響度の大きな研究を除いて解析したメタアナリシスの結果を比較する。

最初に、表 8.4 に、外れ値または影響度が大きいとされた研究を示す。研究 1 は 5 つの指標で外れ値または影響度の大きな研究として検出され、研究 7, 9 は 4 つの指標で検出され、研究 15 は 3 つの指標で検出された。逆に、研究 1, 2, 7, 9, 10, 15, 18 を除く研究は、どの指標でも検出されなかった。

表 8.4 外れ値または影響度が大きいとされた研究

指標	検出された研究
相対距離	研究 2, 7, 15, 18
標準化残差	研究 1, 7, 9, 15, 18
ベイズ流 P 値	研究 1, 9
診断オッズ比	研究 1, 7, 9, 10
SROC 曲線の AUC への影響度	研究 1, 15
尺度混合正規分布における尺度パラメータの事後推定値	研究 1, 7, 9

研究 1, 9, 15 は予め検出されることを予想していた研究であるが、研究 7 も推定値への影響度が大きいことが、提案法の適用によって明らかになった。研究 7 は、SROC 曲線からの乖離は大きくないものの、感度が最も低く、参加者数が多いことにより、推定値への影響度が大きかったと考えられる。

次に、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を表 8.5 および表 8.6 に示す。表 8.5 は、外れ値とされた個々の

研究を除外した結果であり、表 8.6 は、外れ値とされた複数の研究を除外する観点からの感度解析である。外れ値または影響度の大きな研究を除外した結果、感度は 0.41 から 0.48、偽陽性率は 0.18 から 0.24、SROC 曲線の AUC は 0.550 から 0.634 に、診断オッズ比は 2.52 から 3.13 の範囲の値となった。次に、図 8.5 および図 8.6 に感度と偽陽性率の散布図、および SROC 曲線を示す。図 8.6 の相対距離に関しては、感度と偽陽性率の両方が低い研究を集中的に除くことにより、SROC 曲線の形状が大きく変化した。また、研究 1 または 15 の片方のみを除外する場合については SROC 曲線の形状に変化が見られた。上記以外については、SROC 曲線の形状について、特段の影響は認められなかった。

提案法を用いて、外れ値に相当する研究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 AUC の推定値や SROC 曲線の形状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が特定の研究結果に大きく左右され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力診断と感度分析の結果、診断法のメタアナリシスおよびその結論の頑健性が支持される結果となった。

表 8.5 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（感度，偽陽性率，SROC 曲線の AUC，診断オッズ比）

	感度 [95%信頼区間]	偽陽性率 [95%信頼区間]	SROC 曲線の AUC [95%信頼区間]	診断オッズ比 [95%信頼区間]
全ての研究を用いた結果	0.44 [0.33, 0.56]	0.22 [0.13, 0.34]	0.588 [0.474, 0.676]	2.82 [1.75, 4.60]
外れ値または影響度の大きな研究を除いた結果				
研究 1	0.44 [0.33, 0.56]	0.21 [0.12, 0.31]	0.623 [0.523, 0.698]	3.04 [2.06, 4.72]
研究 2	0.46 [0.34, 0.57]	0.24 [0.14, 0.36]	0.589 [0.473, 0.678]	2.70 [1.67, 4.46]
研究 7	0.47 [0.36, 0.57]	0.23 [0.13, 0.36]	0.595 [0.498, 0.671]	3.02 [1.88, 4.95]
研究 9	0.44 [0.32, 0.56]	0.24 [0.14, 0.36]	0.603 [0.509, 0.673]	2.52 [1.71, 3.72]
研究 10	0.42 [0.32, 0.54]	0.22 [0.12, 0.35]	0.575 [0.463, 0.662]	2.67 [1.66, 4.46]
研究 15	0.41 [0.32, 0.51]	0.20 [0.12, 0.29]	0.560 [0.421, 0.666]	2.92 [1.83, 4.82]
研究 18	0.46 [0.35, 0.57]	0.24 [0.13, 0.36]	0.586 [0.474, 0.673]	2.87 [1.74, 4.89]

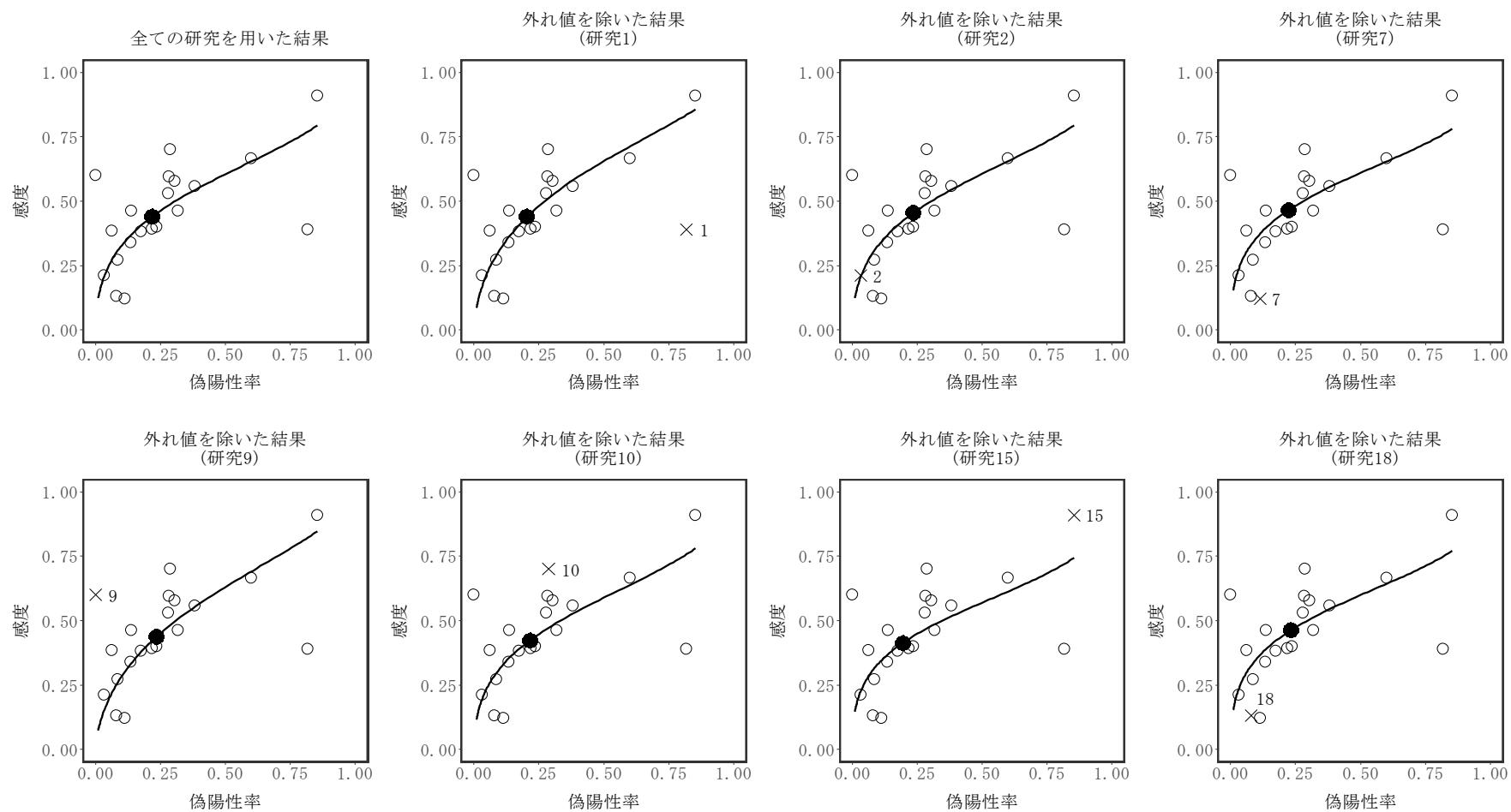


図 8.5 外れ値とされた研究を除外した場合のメタアナリシスの結果 (散布図および SROC 曲線)

(○ : 個別研究の感度および偽陽性率, × : 外れ値とされた研究, ● : メタアナリシスによる全体の感度および偽陽性率)

表 8.6 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（感度，偽陽性率，SROC 曲線の AUC，診断オッズ比）

	外れ値	感度 [95%信頼区間]	偽陽性率 [95%信頼区間]	SROC 曲線の AUC [95%信頼区間]	診断オッズ比 [95%信頼区間]
全ての研究を用いた結果	—	0.44 [0.33, 0.56]	0.22 [0.13, 0.34]	0.588 [0.474, 0.676]	2.82 [1.75, 4.60]
外れ値または影響度の大きな研究を除いた結果					
相対距離	研究 2, 7, 15, 18	0.48 [0.40, 0.56]	0.23 [0.14, 0.35]	0.550 [0.436, 0.646]	3.13 [1.82, 5.66]
標準化残差	研究 1, 7, 9, 15, 18	0.46 [0.36, 0.55]	0.22 [0.14, 0.31]	0.620 [0.537, 0.682]	3.12 [2.24, 4.48]
ベイズ流 P 値	研究 1, 9	0.44 [0.32, 0.57]	0.23 [0.14, 0.34]	0.634 [0.560, 0.693]	2.72 [1.98, 3.80]
診断オッズ比	研究 1, 7, 9, 10	0.44 [0.33, 0.57]	0.23 [0.13, 0.36]	0.618 [0.557, 0.670]	2.79 [2.00, 3.95]
SROC 曲線の AUC への影響度	研究 1, 15	0.41 [0.31, 0.52]	0.18 [0.12, 0.26]	0.606 [0.471, 0.695]	3.13 [2.12, 4.86]
尺度混合正規分布における尺度パラメータの事後推定値	研究 1, 7, 9	0.46 [0.35, 0.58]	0.23 [0.14, 0.35]	0.634 [0.569, 0.688]	2.94 [2.13, 4.13]

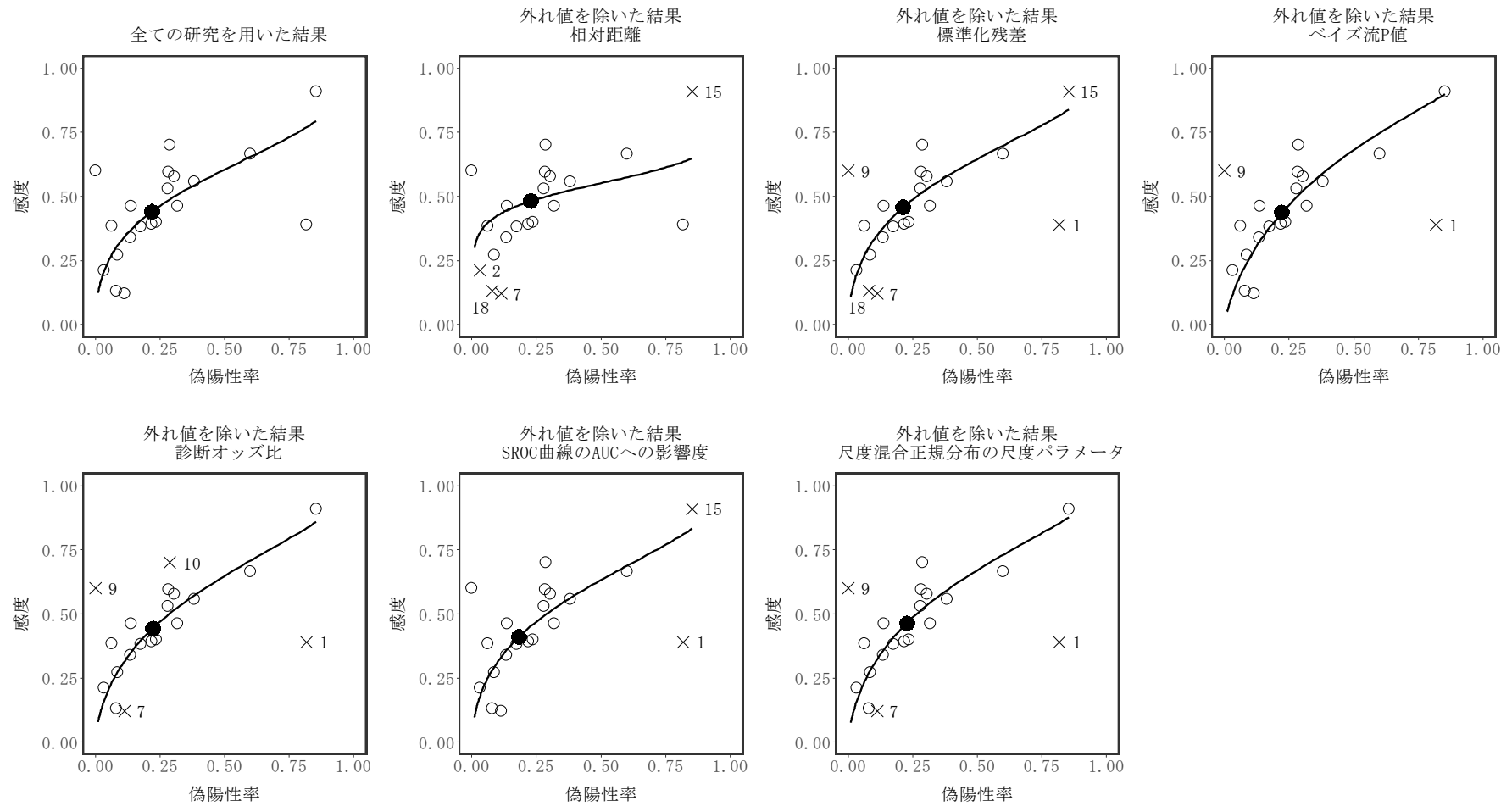


図 8.6 外れ値とされた研究を除外した場合のメタアナリシスの結果（散布図および SROC 曲線）

(○：個別研究の感度および偽陽性率， ×：外れ値とされた研究， ●：メタアナリシスによる全体の感度および偽陽性率)

第9章 実データへの適用（喘息患者における好酸球性気道炎症の代替マーカー）

9.1 適用データの紹介

本章では、3章から7章で提案された指標の例示として、喘息患者における好酸球性気道炎症の代替マーカーの評価のデータに提案法を適用する。この診断法のメタアナリシスは Korevaar et al. [18] によって実施されたものである。好酸球性気道炎症は、喘息患者における副腎皮質ステロイドの反応性と関連が認められているものの、気道からの直接採取法は侵襲性があり、手間もかかる手技である。それゆえに、侵襲性が最小限となるような診断マーカーの探索が続けられている。Korevaar et al. [18]は、診断法のメタアナリシスの方法を用いて、いくつかの診断マーカーのメタアナリシスを実施した。本論文では、いくつかのメタアナリシスのうち、呼気中一酸化窒素濃度のデータに提案法を適用する。この診断法のメタアナリシスには12研究が含まれており、それらの個別研究データは表9.1に示され、感度および偽陽性率の散布図およびSROC曲線は図9.1に示される。

図9.1より、研究3は、偽陽性率が低いにも関わらず、感度が高いことより、外れ値であると考えられる。本事例への適用においては、研究3の検出についての性能確認を行い、また、それ以外にも、結果に強い影響を与える研究の検出についての性能評価を行う。

表 9.1 喘息患者における好酸球性気道炎症の代替マーカーのデータ

番号	研究	TP	FP	FN	TN	感度	偽陽性率
						[95%信頼区間]	[95%信頼区間]
1	Lemiere et al (2006)	17	24	4	15	0.81 [0.58, 0.95]	0.62 [0.45, 0.77]
2	ten Brinke et al (2001)	19	11	6	28	0.76 [0.55, 0.91]	0.28 [0.15, 0.45]
3	Hillas et al (2011)	10	1	4	25	0.71 [0.42, 0.92]	0.04 [0.00, 0.20]
4	Meijer et al (2002)	43	8	36	29	0.54 [0.43, 0.66]	0.22 [0.10, 0.38]
5	Westerhof et al (2014)	92	66	21	147	0.81 [0.73, 0.88]	0.31 [0.25, 0.38]
6	Yap et al (2013)	18	14	3	19	0.86 [0.64, 0.97]	0.42 [0.25, 0.61]
7	De Carvalho-Pinto et al (2012)	34	2	19	12	0.64 [0.50, 0.77]	0.14 [0.02, 0.43]
8	Hastie et al (2013)	49	58	27	104	0.64 [0.53, 0.75]	0.36 [0.28, 0.44]
9	Tseliou et al (2010)	23	1	22	10	0.51 [0.36, 0.66]	0.09 [0.00, 0.41]
10	Jia et al (2012)	17	1	26	12	0.40 [0.25, 0.56]	0.08 [0.00, 0.36]
11	Greulich et al (2012)	48	14	29	44	0.62 [0.51, 0.73]	0.24 [0.14, 0.37]
12	Schleich et al (2013)	147	59	78	224	0.65 [0.59, 0.72]	0.21 [0.16, 0.26]

TP : true positive 真陽性, FP : false positive 偽陽性, FN : false negative 偽陰性,
 TN : true negative 真陰性

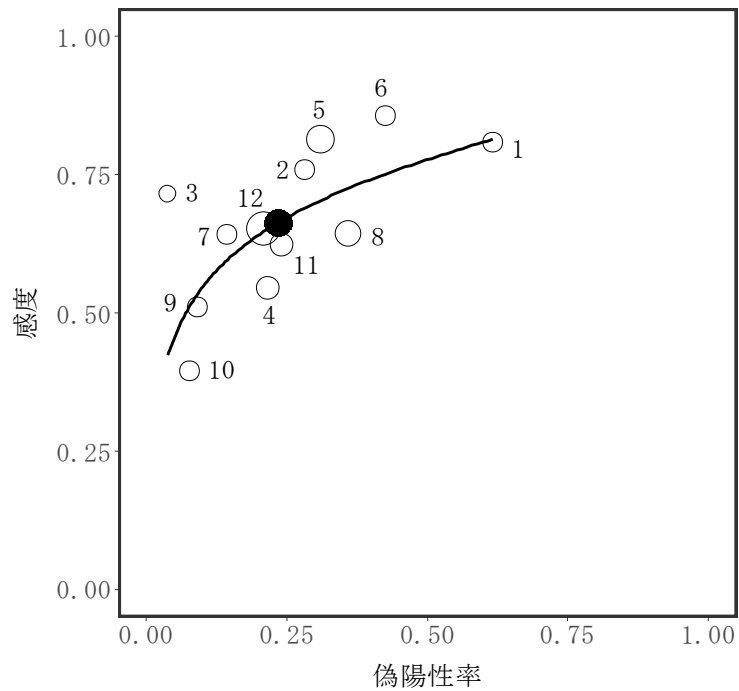


図 9.1 喘息患者における好酸球性気道炎症の代替マーカーのデータ（散布図および SROC 曲線）

- ：個別研究の感度および偽陽性率 [印の大きさは参加者数を反映する]，
- ：診断法のメタアナリシスにより推定された全体の感度および偽陽性率

9.2 提案法の適用結果

相対距離

図 9.2 に、相対距離の結果を示した。図 9.2 の縦軸は相対距離の値であり、横軸は leave-one-out 交差検証法によって除外された研究の番号である。図 9.2 には 10 個のパネルがあり、(a) 感度についての相対距離、(b) 偽陽性率についての相対距離、(c) 統合相対距離、(d) 平均相対距離、(e) 診断オッズ比についての相対距離を示す。

まず、感度と偽陽性率のそれぞれに対する相対距離の結果を述べる。相対距離

(感度)については、絶対値で 0.10 や 0.05 を超える研究はなかったが、高い感度をもつ研究 5、低い感度をもつ研究 10 について、絶対値で 0.025 を超え、他の研究と比較して大きな絶対値となった。相対距離(偽陽性率)については、低い偽陽性率をもつ研究 3 について、0.10 より大きな絶対値となった。8 章の事例と同様であるが、相対距離(感度)と相対距離(偽陽性率)を比べると、全体的に、相対距離(偽陽性率)の方が、絶対値が大きくなっている。これは、相対距離を求める際の分母(感度または偽陽性率の全体の推定値)が異なり、偽陽性率の方が、分母が小さくなるためである。

次に、統合相対距離に関して述べると、0.05 よりも大きな統合相対距離をもつ研究は認められなかったものの、平均相対距離に関しては研究 3、10 が 0.05 よりも大きな値となり、これらの研究は影響力の大きな研究であることが示唆された。研究 3 は最も低い偽陽性率をもつ研究であり、研究 10 は最も低い感度をもつ研究である。統合相対距離と平均相対距離の比較に関しては、どちらの指標に対しても研究 10 が最大の値となり、その次に研究 1、3、5、6 が大きな値となる点で共通しており、顕著な違いは認められなかった。

最後に、診断オッズ比を用いた相対距離の結果を述べる。診断オッズ比に関しては、診断オッズ比に関しては、図 9.2 より、研究 3、8 が 0.10 より大きな絶対値となった。感度が高く偽陽性率が低い場合に診断オッズ比は大きくなり、感度が低く偽陽性率が高い場合に診断オッズ比は小さくなるという、診断オッズ比の性質のため、図 9.1 で、左上または右下方向に外れている研究が検出された。

相対距離の適用結果としては、感度または偽陽性率への影響が大きい研究として研究 3、5、10 が認められ、診断オッズ比への影響度が大きい研究として研究 3 と 8 が認められた。

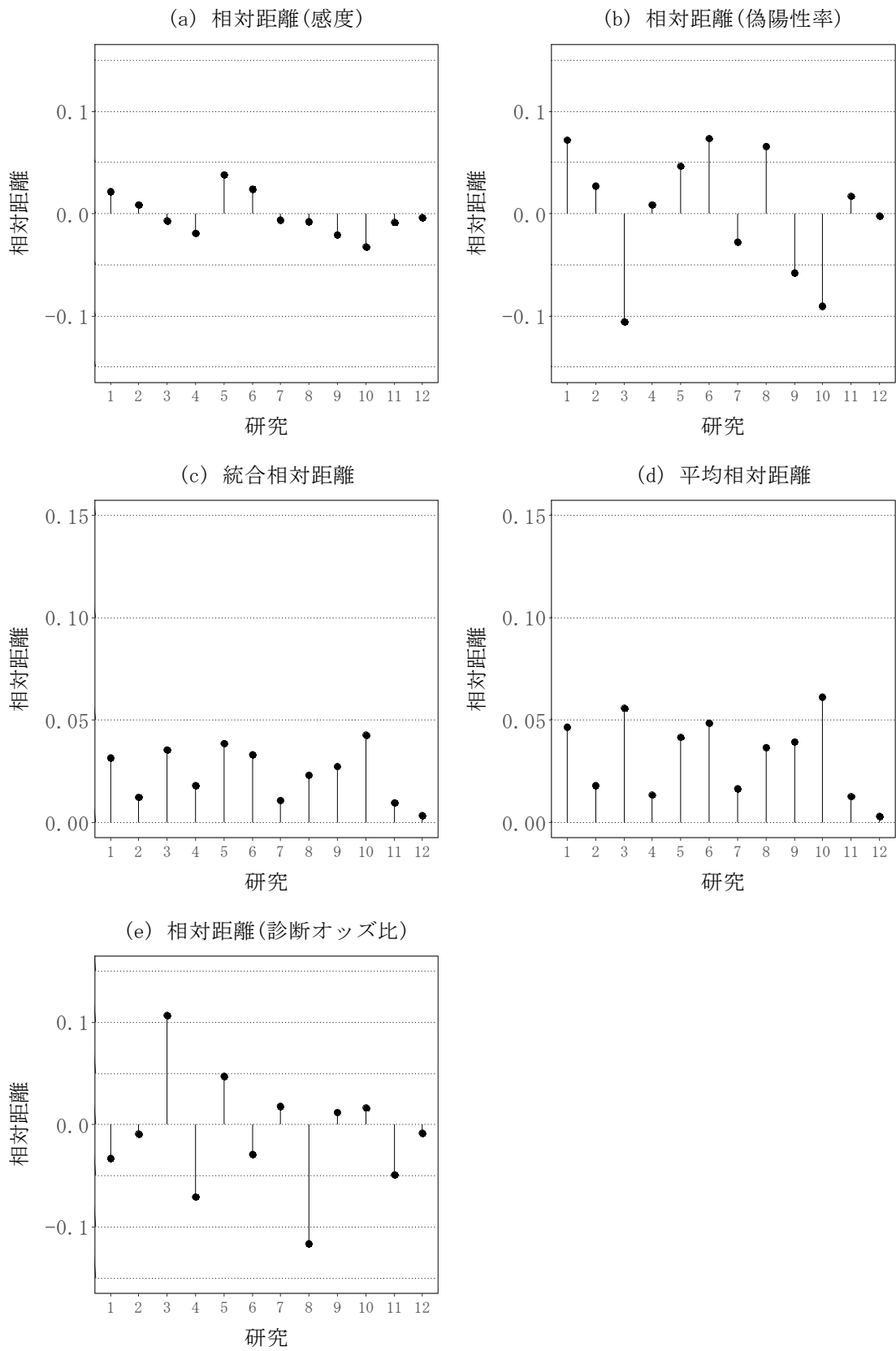


図 9.2 相対距離の結果

標準化残差

図 9.3 に、標準化残差の結果を示した。図 9.3 の縦軸は標準化残差の値であり、横軸は leave-one-out 交差検証法によって除外された研究の番号である。図 9.3 には 10 個のパネルがあり、(a) 感度についての標準化残差、(b) 偽陽性率についての標準化残差、(c) 統合標準化残差、(d) 平均標準化残差、(e) 診断オッズ比についての標準化残差を示す。

まず、感度と偽陽性率のそれぞれに対する標準化残差の結果を述べる。感度については、低い感度をもつ研究 10 について、標準正規分布の上側 5%点/下側 5%点である ± 1.64 の範囲外の値となり、研究 10 が外れ値であることが示唆された。偽陽性率については、高い偽陽性率をもつ研究 1 に対して、また、低い偽陽性率をもつ研究 3 について、 ± 1.64 の範囲外の値となった。

次に、統合標準化残差に関して述べる。閾値としては、自由度 2 のカイ 2 乗分布の上側 10%点である 4.61 を用いた。結果として、最も高い偽陽性率をもつ研究 1 が 4.61 より大きな統合標準化残差をもち、この研究が外れ値であることが示唆された。また、研究 3, 10 も比較的大きな統合標準化残差を示したものの、閾値の 4.61 より小さな値であった。平均標準化残差に関しては、統合標準化残差と同じく、研究 1 について大きな平均標準化残差となった。しかしながら、外れ値と考えていた研究 3 については大きな平均標準化残差が得られなかった。9 章の事例においても、8 章と同様に、外れ値の検出に関して、平均標準化残差よりも統合標準化残差の方が適切な指標と考えられる結果となった。

診断オッズ比に関しては、図 9.3 より、研究 3 の標準化残差が ± 1.64 の範囲外の値となった。これまでの傾向と同様に、プロット上で、左上または右下方向に外れている研究に対して、標準化残差の絶対値が大きくなる傾向がみられた。

標準化残差の適用結果としては、外れ値の可能性のある研究として、研究 1, 3, 10 が認められた。

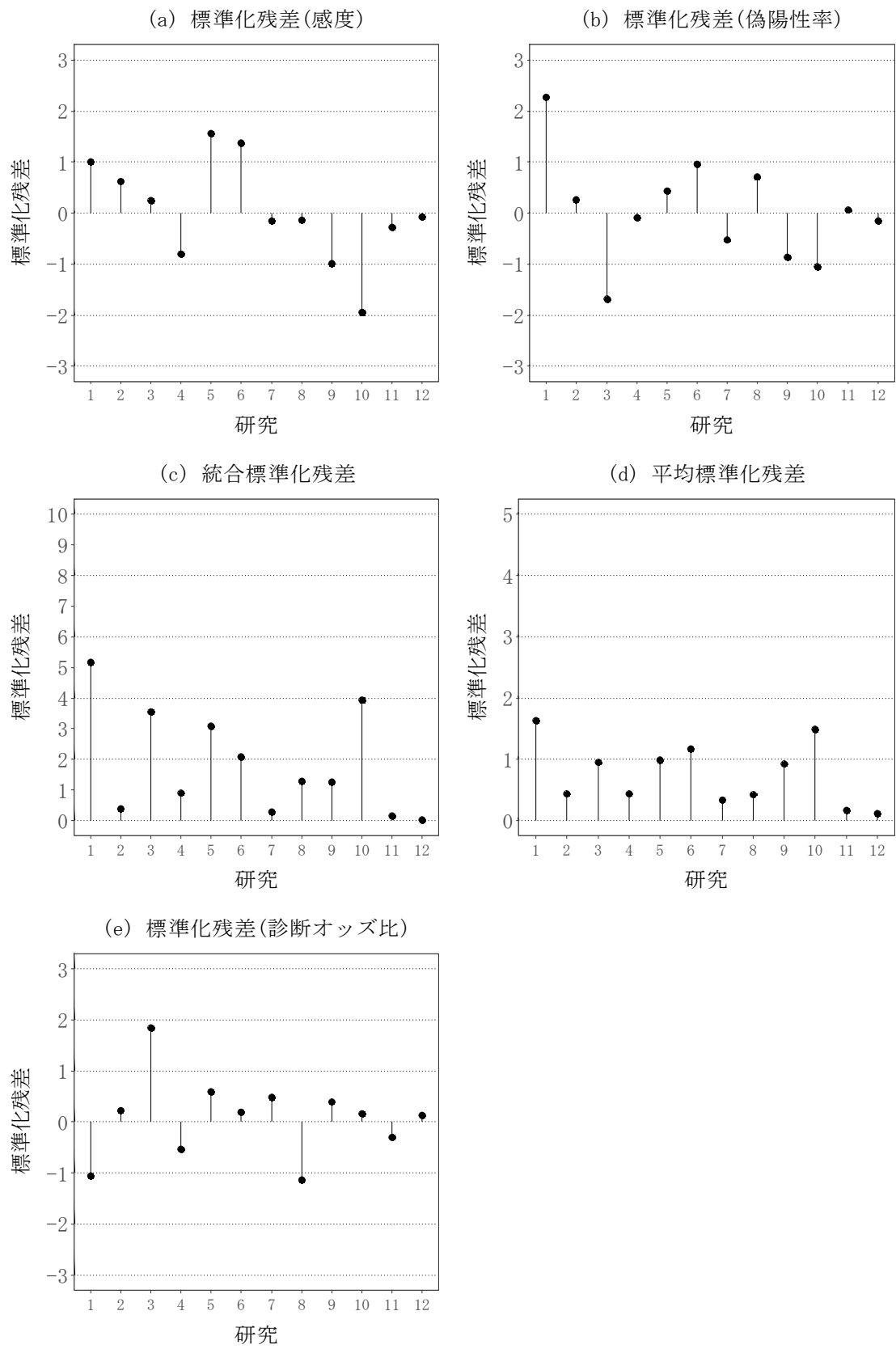


図 9.3 標準化残差の結果

ベイズ流 P 値

表 9.2 にベイズ流 P 値（感度，偽陽性率，感度および偽陽性率を統合した指標，感度および偽陽性率の平均に相当する指標，診断オッズ比）を示す。研究 3 は，偽陽性率，統合した指標，平均に相当する指標および診断オッズ比について，0.15 を下回るベイズ流 P 値となり，診断オッズ比については < 0.05 となった。逆に，研究 3, 6 を除くと，0.3 よりも小さなベイズ流 P 値は全く認められず，ベイズ流 P 値が小さい研究と大きい研究に明確に分かれる結果となった。結果として，研究 3 はベイズ流 P 値の点からは，外れ値の可能性があると判定された。

感度および偽陽性率を統合した指標と，感度および偽陽性率の平均に相当する指標の比較結果について述べる。外れ値と考えられる研究 3 に対しては，統合した指標のベイズ流 P 値は 0.0600，平均に相当する指標のベイズ流 P 値は 0.1049 であり，統合した指標を用いる方が小さくなった。また，仮に，外れ値を検出するための閾値を 0.10 とすると，平均に相当する指標では，研究 3 は外れ値と検出されない結果となった。上記より，感度および偽陽性率の平均に相当する指標よりも，感度および偽陽性率を統合した指標の方が，外れ値の検出に関して，高い性能を示すと考えられた。

表 9.2 ベイズ流 P 値の結果

番号	研究	感度	偽陽性率	統合指標	平均	診断 オッズ比
1	Lemiere et al (2006)	0.9939	0.4712	0.7677	0.7707	0.6390
2	ten Brinke et al (2001)	0.5315	0.9864	0.8004	0.8057	0.5826
3	Hillas et al (2011)	0.2477	0.0553	0.0600	0.1049	0.0198
4	Meijer et al (2002)	0.5920	0.7369	0.7834	0.8130	0.5214
5	Westerhof et al (2014)	0.4142	0.8733	0.7119	0.7135	0.4474
6	Yap et al (2013)	0.2151	0.8048	0.4632	0.4476	0.3505
7	De Carvalho-Pinto et al (2012)	0.9256	0.5526	0.8136	0.8339	0.5117
8	Hastie et al (2013)	0.5681	0.6988	0.7857	0.7904	0.4920
9	Tseliou et al (2010)	0.6874	0.5080	0.7831	0.7282	0.6079
10	Jia et al (2012)	0.4307	0.6892	0.7000	0.6535	0.9619
11	Greulich et al (2012)	0.8058	0.8349	0.9356	0.9447	0.7225
12	Schleich et al (2013)	0.9759	0.9491	0.9975	0.9976	0.9465

SROC 曲線の AUC への影響度

図 9.4 に、各研究をメタアナリシスから除外することによる SROC 曲線の AUC への影響度解析の結果を示す。 ΔAUC_i は 12 研究のほとんど対して、絶対値で 0.01 未満の値となったが、研究 5, 6, 8 については、それぞれ、0.028, 0.017, -0.019 の値となった。これらの 3 つの研究は ΔAUC_i の絶対値が大きいことにより、影響力が大きいことが示唆された。研究 5, 6 は高い感度もつ研究であり、研究 5, 8 は参加者数が多い研究である。このため、これらの研究の AUC への影響度が大きくなったと考えられる。

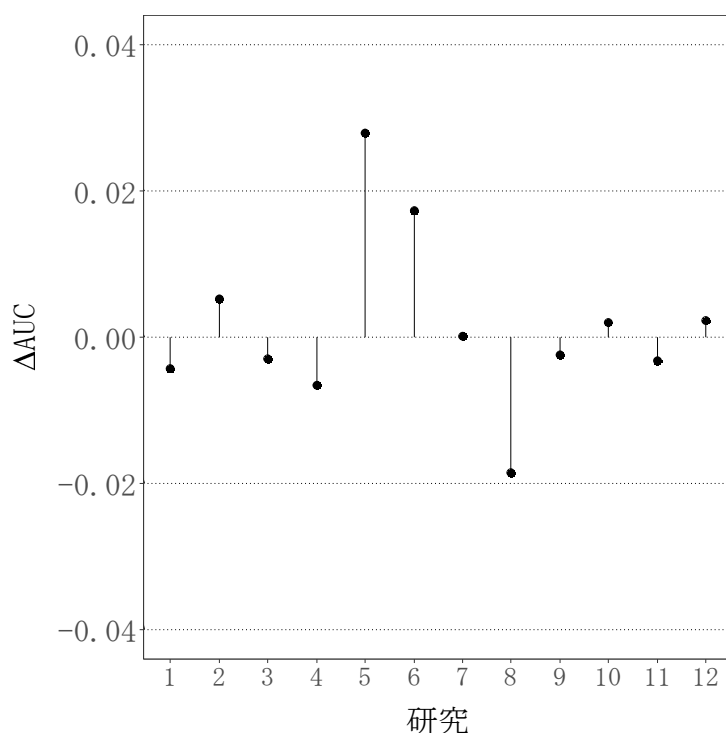


図 9.4 SROC 曲線の AUC への影響度の結果

尺度混合正規分布における尺度パラメータの事後推定値

表 9.3 に尺度混合正規分布における尺度パラメータの事後推定値に関する結果を示す。この表では、感度、偽陽性率および CSP モデルについて、尺度パラメータが 1 よりも大きくなる事後確率を示す。8.3 節と同様に、 λ_i の事前分布として、逆ガンマ分布 $IG(1, 1)$ および指数分布 $EXP(2)$ を用いた。また、本適用においては、外れ値を積極的に発見するための値として、閾値は 0.7 と設定した。

適用結果について述べる。感度については、研究 5 と 10 について $\Pr(\lambda_i > 1) > 0.7$ となった。偽陽性率については、研究 1 と 3 について $\Pr(\lambda_i > 1) > 0.7$ となった。最後に、CSP モデルについては、研究 1, 3, 5 について $\Pr(\lambda_i > 1) > 0.7$ となった。感度の評価で外れ値と検出された研究 10 については、CSP モデルでは検出されない結果となった。

なお、この適用事例においても、 λ_i の事前分布の選択は $\Pr(\lambda_i > 1)$ の値に対

してわずかに影響を与えたものの、外れ値として検出される研究はどちらの事前分布を用いても同じ結果であった。

尺度混合正規分布における尺度パラメータの事後推定値の適用結果としては、外れ値の可能性のある研究として、研究 1, 3, 5, 10 が認められた。

表 9.3 尺度混合正規分布における尺度パラメータの事後推定値（尺度パラメータが 1 よりも大きくなる事後確率）

番号	研究	感度		偽陽性率		CSP モデル	
		IG	EXP	IG	EXP	IG	EXP
1	Lemiere et al (2006)	0.5806	0.5497	0.8055	0.7655	0.7422	0.7279
2	ten Brinke et al (2001)	0.5720	0.5718	0.5488	0.5342	0.5155	0.5063
3	Hillas et al (2011)	0.5442	0.5386	0.8332	0.8490	0.7830	0.8228
4	Meijer et al (2002)	0.6337	0.6179	0.5448	0.5419	0.5555	0.5543
5	Westerhof et al (2014)	0.7377	0.7150	0.6069	0.5843	0.7319	0.7167
6	Yap et al (2013)	0.6647	0.6554	0.6221	0.5983	0.6539	0.6558
7	De Carvalho-Pinto et al (2012)	0.5538	0.5441	0.5665	0.5653	0.5025	0.5149
8	Hastie et al (2013)	0.5884	0.5827	0.6600	0.6332	0.6059	0.6104
9	Tseliou et al (2010)	0.6282	0.6048	0.5812	0.5787	0.5741	0.5754
10	Jia et al (2012)	0.7429	0.7114	0.6033	0.5875	0.6992	0.6877
11	Greulich et al (2012)	0.5659	0.5551	0.5528	0.5488	0.4974	0.4878
12	Schleich et al (2013)	0.5897	0.5811	0.6374	0.6118	0.5943	0.5886

IG : λ_i の事前分布として逆ガンマ分布を用いる場合, EXP : λ_i の事前分布として指数分布を用いる場合

9.3 メタアナリシスの結果に対する影響力の評価

最後に、全ての 12 研究を用いたメタアナリシスと、外れ値または影響度の大きな研究を除いて解析したメタアナリシスの結果を比較する。最初に、表 9.4 に外れ値または影響度が大きいとされた研究を示す。

表 9.4 外れ値または影響度が大きいとされた研究

指標	検出された研究
相対距離	研究 3, 5, 10
標準化残差	研究 1, 3, 10
ベイズ流 P 値	研究 3
診断オッズ比	研究 3, 8
SROC 曲線の AUC への影響度	研究 5, 6, 8
尺度混合正規分布における尺度パラメータの事後推定値	研究 1, 3, 5, 10

研究 3 は 5 つの指標で外れ値または影響度の大きな研究として検出され、研究 5, 10 は 3 つの指標で検出され、研究 1, 8 は 2 つの指標で検出された。逆に、研究 2, 4, 7, 9, 11, 12 は、どの指標でも検出されなかった。研究 3 は予め検出されることを予想していた研究であるが、研究 1, 5, 8, 10 についても推定値への影響度が大きいことが、提案法の適用によって明らかになった。研究 5 と 8 は、SROC 曲線からの乖離は大きくないものの、参加者数が多いことにより、外れ値として検出された。また、研究 1 と 10 は、SROC 曲線からの乖離は大きくないものの、感度または偽陽性率が極端であることより、外れ値として検出された。

次に、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を表 9.5 および表 9.6 に示す。表 9.5 は、外れ値とされた個々の研究を除外した結果であり、表 9.6 は、外れ値とされた複数の研究を除外する観点からの感度解析である。外れ値または影響度の大きな研究を除外した結果、感度は 0.62 から 0.68、偽陽性率は 0.19 から 0.29、SROC 曲線の AUC は 0.680 から 0.762 に、診断オッズ比は 5.03 から 7.19 の範囲の値となった。次に、図 9.5 および図 9.6 に感度と偽陽性率の散布図、および SROC 曲線を示す。図 9.5 より、研究 5, 6, 10 を除外することにより、SROC 曲線は下側に押し下げられることが分かり、図 9.6 より、研究 5, 6, 10 のうちの複数の研究をメタアナリシスから除外する場合については、SROC 曲線の印象が変わる結果となった。

提案法を用いて詳細に評価した結果、研究 3 が外れ値に相当する可能性があり、研究 5, 6, 10 は SROC 曲線の形状、および、AUC への影響が大きいことが分かった。また、複数の研究をメタアナリシスから除外する場合については、AUC の推定値が 0.05 程度、変化する結果となった。しかしながら、特定の研究を除外するのみで全体の結論が変わるような影響力は認められず、診断法のメタアナリシスの結果と、その結論の妥当性が確認されたということができよう。

表 9.5 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（感度，偽陽性率，SROC 曲線の AUC，診断オッズ比）

	感度 [95%信頼区間]	偽陽性率 [95%信頼区間]	SROC 曲線の AUC [95%信頼区間]	診断オッズ比 [95%信頼区間]
全ての研究を用いた結果	0.66 [0.57, 0.75]	0.24 [0.13, 0.35]	0.742 [0.645, 0.804]	6.44 [3.90, 11.40]
外れ値または影響度の大きな研究を除いた結果				
研究 1	0.65 [0.55, 0.74]	0.22 [0.13, 0.31]	0.741 [0.597, 0.811]	6.65 [4.11, 11.72]
研究 3	0.67 [0.56, 0.77]	0.26 [0.16, 0.37]	0.744 [0.654, 0.801]	5.75 [3.58, 9.34]
研究 5	0.64 [0.55, 0.73]	0.23 [0.12, 0.36]	0.715 [0.619, 0.788]	6.14 [3.47, 11.88]
研究 6	0.65 [0.55, 0.74]	0.22 [0.12, 0.34]	0.724 [0.606, 0.800]	6.63 [3.74, 12.85]
研究 8	0.67 [0.55, 0.77]	0.22 [0.12, 0.35]	0.759 [0.654, 0.822]	7.19 [4.22, 12.90]
研究 10	0.68 [0.60, 0.77]	0.26 [0.15, 0.38]	0.738 [0.635, 0.805]	6.34 [3.74, 11.46]

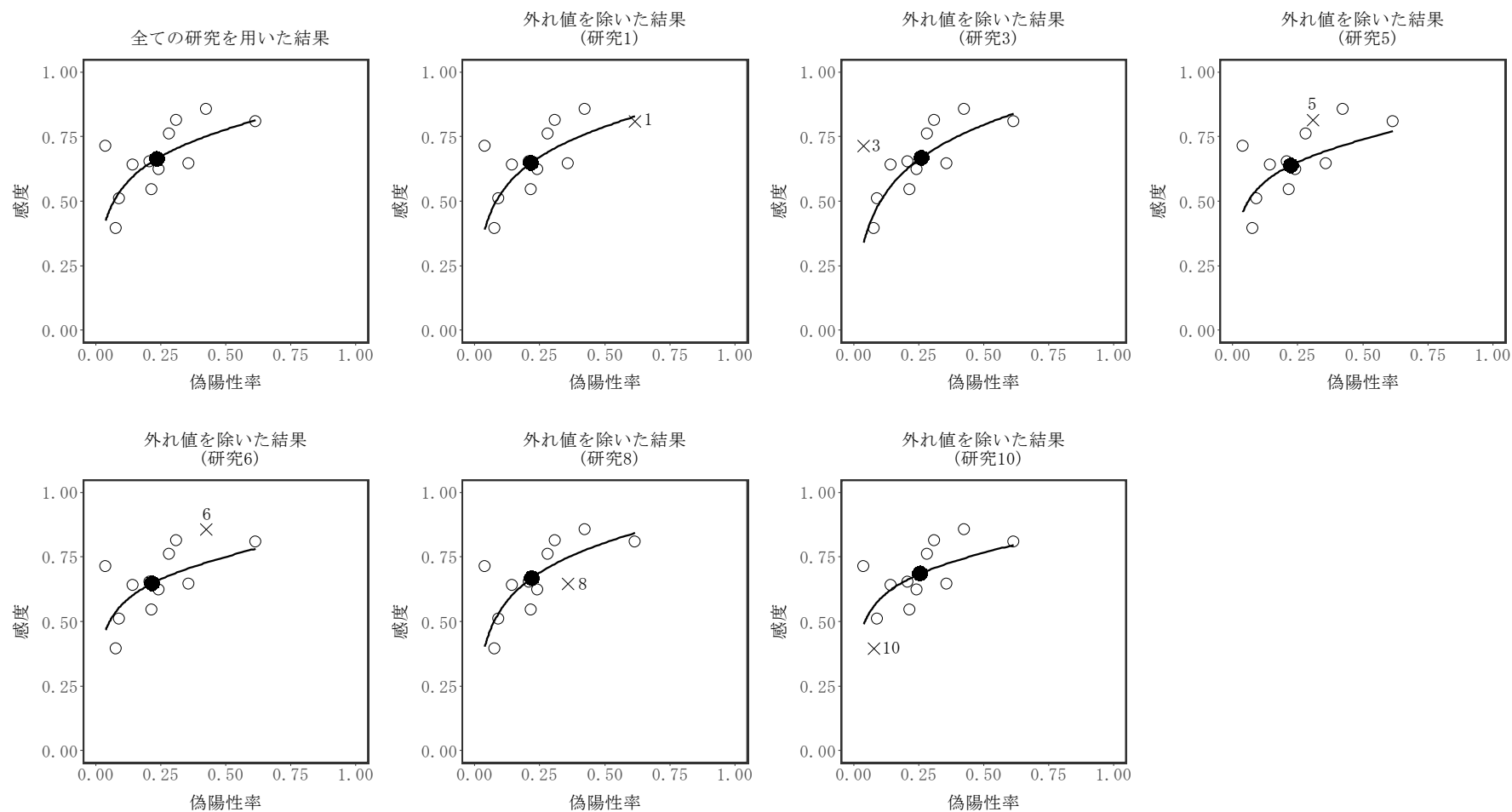


図 9.5 外れ値とされた研究を除外した場合のメタアナリシスの結果 (散布図および SROC 曲線)

(○ : 個別研究の感度および偽陽性率, × : 外れ値とされた研究, ● : メタアナリシスによる全体の感度および偽陽性率)

表 9.6 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（感度，偽陽性率，SROC 曲線の AUC，診断オッズ比）

	外れ値	感度 [95%信頼区間]	偽陽性率 [95%信頼区間]	SROC 曲線の AUC [95%信頼区間]	診断オッズ比 [95%信頼区間]
全ての研究を用いた結果	—	0.66 [0.57, 0.75]	0.24 [0.13, 0.35]	0.742 [0.645, 0.804]	6.44 [3.90, 11.40]
外れ値または影響度の大きな研究					
相対距離	研究 3, 5, 10	0.66 [0.58, 0.75]	0.29 [0.17, 0.42]	0.705 [0.613, 0.782]	5.03 [2.87, 8.94]
標準化残差	研究 1, 3, 10	0.67 [0.57, 0.77]	0.26 [0.18, 0.34]	0.733 [0.541, 0.806]	5.87 [3.65, 9.54]
ベイズ流 P 値	研究 3	0.67 [0.56, 0.77]	0.26 [0.16, 0.37]	0.744 [0.654, 0.801]	5.75 [3.58, 9.34]
診断オッズ比	研究 3, 8	0.67 [0.55, 0.79]	0.25 [0.14, 0.37]	0.762 [0.670, 0.816]	6.41 [3.86, 10.50]
SROC 曲線の AUC への影響度	研究 5, 6, 8	0.62 [0.52, 0.72]	0.19 [0.08, 0.35]	0.694 [0.557, 0.796]	7.15 [3.32, 17.44]
尺度混合正規分布における尺度パラメータの事後推定値	研究 1, 3, 5, 10	0.64 [0.56, 0.73]	0.26 [0.17, 0.35]	0.680 [0.491, 0.787]	5.31 [3.22, 9.14]

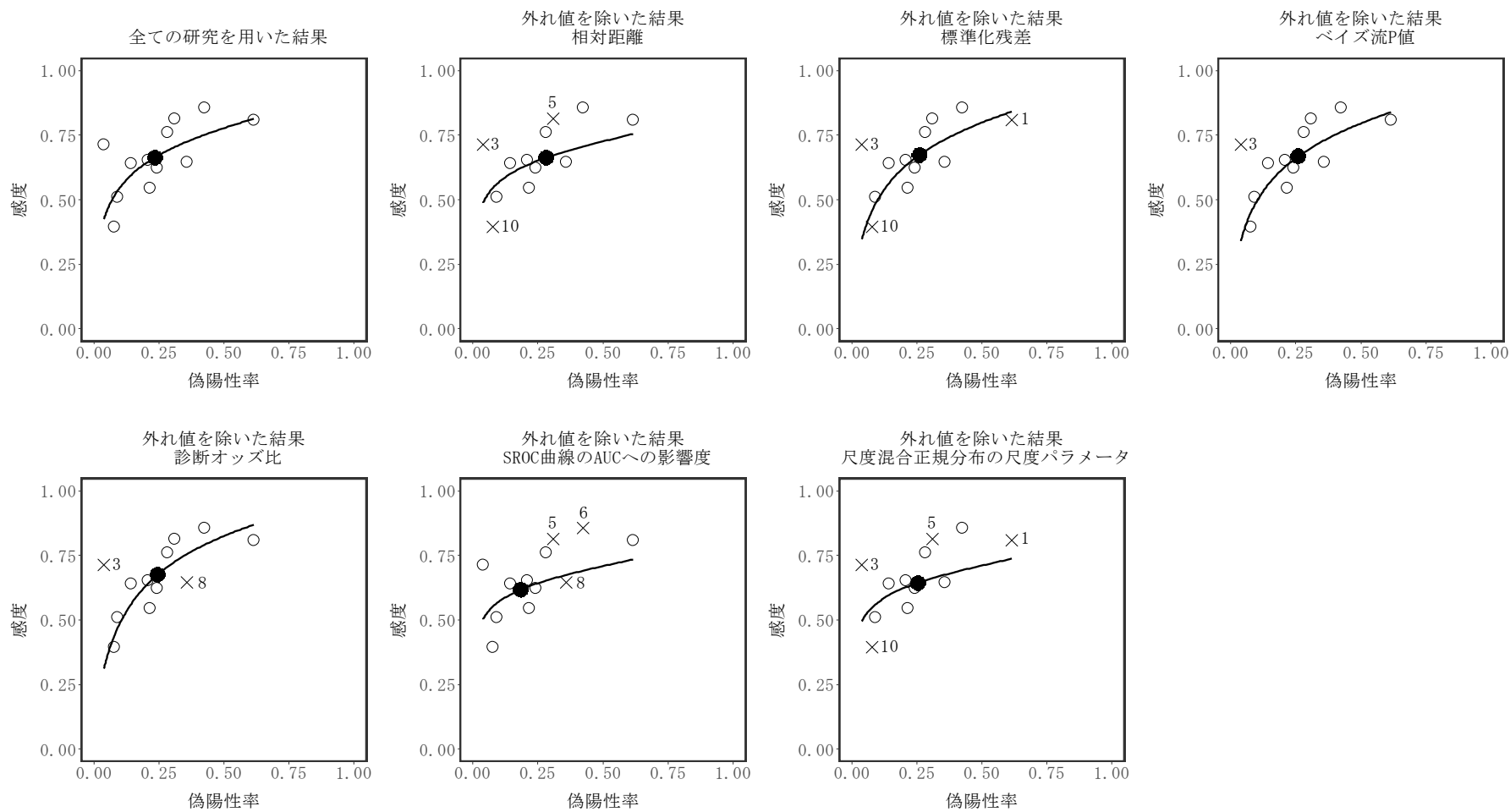


図 9.6 外れ値とされた研究を除外した場合のメタアナリシスの結果（散布図および SROC 曲線）

（○：個別研究の感度および偽陽性率， ×：外れ値とされた研究， ●：メタアナリシスによる全体の感度および偽陽性率）

第 10 章 考察

10.1 研究成果のまとめ

本論文では、ベイズ統計学を用いた、診断法のメタアナリシスにおける外れ値の検出および影響力診断の方法を提案した。同様の方法はネットワークメタアナリシスについて既に議論されているものの[15]、診断法のメタアナリシスでは、感度および偽陽性率を 2 変量モデルによって評価する必要がある、診断オッズ比や SROC 曲線の AUC のような独特な要約指標が用いられるという特徴がある。従って、外れ値の検出および影響力診断について、診断法のメタアナリシスのための特別な方法が必要である。診断法のメタアナリシスにおいては、フォレストプロットや、SROC 曲線とともに描かれる感度と偽陽性率の散布図を用いて、外れ値または影響度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、本論文では定量的な評価方法を提案した。なお、本論文ではベイズ流の外れ値の検出および影響力診断の方法として提案を行ったが、相対距離、標準化残差および SROC 曲線の AUC への影響度については、頻度論に基づく手法にも同様に適用することができる。

実データへの適用（8 章，9 章）では、小児の尿路感染症患者における膀胱尿管逆流の超音波によるスクリーニング、および、喘息患者における好酸球性気道炎症の代替マーカーについての診断法のメタアナリシスへの適用を通して、提案法の有用性を示した。いずれの事例においても、提案法は外れ値または影響度の大きな研究を検出することができ、全体の統合解析の結果への影響度を定量的に評価することができた。さらに、提案法は、感度または偽陽性率が極端であるが SROC 曲線からの乖離が大きくない研究（9.3 節）や、参加者数が多いために推定値への影響が大きい研究（8.4 節，9.3 節）に対しても、外れ値として検出

する性能を示した。これらの研究は、SROC 曲線から大きく外れている研究に比べて、プロットによる目視では見逃す可能性があるため、提案法による外れ値の検出は有用であると考えられる。

適用事例においては、全体の結論を変えるような影響力の大きい研究は見つからなかったが、影響度の定量的な評価と感度分析によって、元々のメタアナリシスの結論の頑健性は支持された。このような定量的な評価は、診断法のメタアナリシスから統合されたエビデンスや、それに基づく政策立案および医療技術評価に有用であると考えられる。

10.2 各提案法の特徴

本論文では、診断法のメタアナリシスにおける外れ値の検出のための方法として、相対距離、標準化残差、ベイズ流 P 値、SROC 曲線の AUC への影響度、尺度混合正規分布における尺度パラメータの事後推定値の 5 つの指標を提案した。また、感度と偽陽性率の一方を評価する指標と、感度と偽陽性率を同時に評価する指標について提案した。本節では、各指標の特性について述べる。

まず、本論文では 5 つの指標を提案したが、これらの提案法については、それぞれ異なった観点から外れ値を検出する指標であり、特定の指標が常に優れることはなく、それぞれの指標の特性に合致するような外れ値が検出されるものとする。それぞれの指標の特性については、3 章から 7 章に説明があるものの、改めて、表 10.1 に整理する。

表 10.1 外れ値を検出するための指標とその特性

指標	指標の特性
相対距離	leave-one-out 交差検証法によって、感度や偽陽性率の推定値への影響の大きさを評価する指標である。
標準化残差	leave-one-out 交差検証法によって、観測値についての、leave-one-out データセットから得られる事後予測平均からの残差を、事後予測分散で標準化した指標である。
ベイズ流 P 値	事後予測分布を用いたモデル確認の方法であり、観測値と、想定されるモデル（事後予測サンプル）の乖離を評価する指標である。
SROC 曲線の AUC への影響度	leave-one-out 交差検証法によって、SROC 曲線の AUC の推定値への影響を評価する指標である。
尺度混合正規分布の尺度パラメータの事後推定値	誤差の分布に尺度パラメータを導入することにより、誤差の分布を、裾が重く、外れ値を許容する分布に拡張することができる。極端な観測値に対しては、尺度パラメータの推定値が大きくなる関係を用いて、外れ値の検出を行う指標である。

次に、感度または偽陽性率の一方を評価する指標と、感度と偽陽性率を同時に評価する指標について述べる。本論文では、診断法のメタアナリシスに対して、2変量モデルを用いる解析を行っている。従って、推定結果への影響力の評価においても、1変量ずつの評価のみではなく、2変量を同時に評価することが重要であり、本論文では、2変量の影響度を同時に評価する指標を開発した。しかしながら、診断法のメタアナリシスにおいては、感度と偽陽性率のそれぞれに対する評価も重要であり、感度または偽陽性率の周辺的な指標を用いることにより、影響度の評価が可能となると考える。また、実データへの適用においては、感度と偽陽性率の一方については外れているものの、他方については平均的な値であり、2変量を統合することによって、外れ値として検出されにくくなった場合も存在した。それゆえに、2変量を統合した指標と、周辺的な1変量の指標は補完的に用いることができると考える。

また、感度と偽陽性率を同時に評価する指標に関して、本論文では、標準化残差とベイズ流 P 値に対して、感度と偽陽性率の間の相関関係を考慮した指標を提案した。これらの相関を考慮した指標は、実データへの適用において、相関関係を考慮しない平均値を用いた指標よりも、外れ値と考えられる研究の検出に関して高い性能を示すことが確認された。

最後に、事例への適用結果においては、複数の提案法で、外れ値と判定される場合が認められた。このような場合は、外れ値の度合いがより確からしいと考えられるため、複数の指標を適用し、異なった観点から外れ値の可能性を検討することは有用と考える。

10.3 適用上の留意点

本節では、提案法を実データへ適用するにあたっての留意点を3点挙げる。

1 点目は閾値の設定に関する点である。提案法においては、外れ値を検出するために、何らかの閾値の設定が必要となる。本論文の事例適用においては、閾値として、標準化残差に対しては標準正規分布の上側 5%点/下側 5%点や、自由度 2 のカイ 2 乗分布の上側 10%点を用い、ベイズ流 P 値に対しては 0.15 や 0.05 を用いた。しかしながら、相対距離や AUC への影響度については統計的な見地からの閾値は設定できなかった。また、メタアナリシスに用いられる研究の数が少ない場合には、各研究への重みが大きくなり、leave-one-out 交差検証法に基づく各指標（具体的には、相対距離、標準化残差、SROC 曲線の AUC への影響度）は比較的大きな値を取りやすくなる。このように、現実のデータへの適応においては、一律の閾値を設定することは困難であるものの、この問題は、外れ値の検出と影響力診断においてしばしば見られる一般的な問題である[15, 26]。適切な閾値は適用事例ごとのケースバイケースで決まり、閾値についての一律の設定方法を定めることは難しいと考える。

なお、AUC の影響度に対する閾値の客観的な設定方法については、頻度論の枠組みにはなるものの、ブートストラップ法を適用することにより、 ΔAUC_i のブートストラップ分布のパーセント点（例えば、2.5%および 97.5%点）を閾値とする方法を提案した[27]。ブートストラップ法を用いることにより、 ΔAUC_i の不確実性についての定量的な評価が可能となり、影響力の大きな研究を検出するための閾値について、客観的かつ合理的な設定が可能となると考える。ブートストラップ法による、影響力の大きな研究の検出については、R パッケージの `dmetatools` で実行することができる[27]。

2 点目としては、本研究では、診断法のメタアナリシスに対して標準的と考えられる 2 変量の変量効果モデルを用いたが、このモデルで用いられるロジット変換は、0 または 1 に近い値における小さな差異を過大評価する可能性がある。この過大評価は、標準化残差、ベイズ流 P 値、尺度混合正規分布における尺度パ

ラメータの事後推定値に対して留意が必要となる。

3点目は外れ値への対応方法に関する点である。まず、今回の提案法による外れ値の検出と影響力診断には多くの解析を伴うため、偶然にも外れ値と判定されてしまう可能性について考慮する必要がある。提案法は外れ値の可能性のある研究を検出するものであり、提案法によって検出された研究について、それらの全てを外れ値であると結論づけるべきではない。また、適用事例においては、元々の結果の頑健さを評価するために、外れ値または影響度の大きな研究を除外してメタアナリシスを行ったが、外れ値または影響度の大きな研究を機械的に主解析から除くことは適切ではない点に留意が必要である。影響力の大きさは研究を主解析から除外する理由にはならず、一方で、外れ値にあたる研究については、その研究デザイン、参加者の特性、地域、施設、診断マーカーのカットオフ値、アウトカムの定義などの点から、外れ値となった原因を詳細に調べる必要がある。

付録 解析プログラム

付録 A 階層ベイズモデルによる診断法のメタアナリシス

【Rプログラム】

```
#### Prepare environment
library("BRugs")
library("R2OpenBUGS")

set.seed(111)
n_iter <- 120000          # Number of iteration in OpenBUGS
n_burnin <- 20000        # Number of burn-in in OpenBUGS
n_thin <- 1              # Thinning interval in OpenBUGS
n_mcmc <- n_iter - n_burnin # Number of samples in OpenBUGS

logit <- function(x) log(x / (1 - x))    # Define logit function
logit.inv <- function(x) (exp(x) / (1 + exp(x))) # Define inverse logit
function

#### Data from Korevaar et al., The Lancet Respiratory Medicine, 2015.
TP <- c(17, 19, 10, 43, 92, 18, 34, 49, 23, 17, 48, 147)
FP <- c(24, 11, 1, 8, 66, 14, 2, 58, 1, 1, 14, 59)
FN <- c(4, 6, 4, 36, 21, 3, 19, 27, 22, 26, 29, 78)
TN <- c(15, 28, 25, 29, 147, 19, 12, 104, 10, 12, 44, 224)
dt1 <- data.frame(TP, FP, FN, TN)

#### Dimercaptosuccinic acid scan or ultrasound in screening for
vesicoureteral reflux among children with urinary tract infections
#### https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD010657.
pub2/full
TP <- c(7, 8, 26, 6, 5, 8, 14, 18, 9, 47, 72, 25, 14, 18, 20, 29, 60,
5, 9, 40)
FP <- c(9, 2, 31, 7, 2, 4, 21, 21, 0, 44, 141, 55, 30, 20, 41, 91, 70,
4, 14, 96)
FN <- c(11, 30, 19, 7, 8, 13, 101, 35, 6, 20, 49, 29, 21, 48, 2, 23, 93,
```

```

33, 8, 20)
TN <- c(2, 60, 71, 44, 30, 19, 165, 135, 37, 109, 356, 118, 97, 214, 7,
147, 250, 46, 36, 64)
dt1 <- data.frame(TP, FP, FN, TN)

### Rename variables
tp <- TP
fp <- FP
n1 <- TP + FN
n2 <- TN + FP
nn <- n1 + n2
n_studies <- length(TP)

### Make the dataset for OpenBUGS
bugs.data(list("tp" = tp, "fp" = fp, "n1" = n1, "n2" = n2,
              "n_studies" = n_studies), data.file = "dtadata.txt")

### Call OpenBUGS
initv0 <- list(mu0 = c(0, 0), sigma1 = 1, sigma2 = 1, rho = 0)

bugs0 <- bugs(data = "dtadata.txt",
              inits = list(initv0),
              parameters.to.save = c("tpr_pool", "fpr_pool", "mu0",
                                     "sigma1", "sigma2", "rho", "ldor_pool", "ltpr",
                                     "lfpr", "ldor"),
              model.file = "Reitsma.txt",
              n.chains = 1,
              n.iter = n_iter,
              n.burnin = n_burnin,
              n.thin = n_thin,
              working.directory = pathname,
              debug = FALSE)

```

【OpenBUGS プログラム】

```
model {
  for(i in 1:n_studies) {
    # Binomial distribution for each study
    tpr[i] ~ dbin(tpr[i], n1[i])
    fpr[i] ~ dbin(fpr[i], n2[i])

    # Logit transformation of sensitivity and false positive rate
    tpr[i] <- exp(m[i, 1]) / (1 + exp(m[i, 1]))
    fpr[i] <- exp(m[i, 2]) / (1 + exp(m[i, 2]))

    # Bivariate normal distribution
    m[i, 1:2] ~ dnorm(mu0[1:2], sigma.inv[1:2, 1:2])

    # logit(sensitivity), logit(FPR) and log(DOR) for each study
    ltp[i] <- m[i, 1]
    lfpr[i] <- m[i, 2]
    ldor[i] <- m[i, 1] - m[i, 2]
  }

  # Priors for means
  mu0[1] ~ dnorm(0, 0.01)
  mu0[2] ~ dnorm(0, 0.01)

  # Priors for the elements of the covariance matrix
  sigma.inv[1:2, 1:2] <- inverse(sigma[1:2, 1:2])
  sigma[1, 1] <- sigma1 * sigma1
  sigma[1, 2] <- sigma1 * sigma2 * rho
  sigma[2, 1] <- sigma1 * sigma2 * rho
  sigma[2, 2] <- sigma2 * sigma2
  sigma1 ~ dunif(0, 10)
  sigma2 ~ dunif(0, 10)
  rho ~ dunif(-1, 1)

  # Pooled summaries
```

```
tpr_pool <- exp(mu0[1]) / (1 + exp(mu0[1]))  
fpr_pool <- exp(mu0[2]) / (1 + exp(mu0[2]))  
ldor_pool <- mu0[1] - mu0[2]  
}
```

付録 B AUC の算出

【R プログラム】

```
# Prepare empty data
fauc1 <- numeric(n_studies)
aucd <- numeric(n_studies)

# Calculate SROC curve
fsroc <- function(theta_tp, theta_fp, sigma_tp, sigma_fp, rho) {
  xpred <- numeric(101)
  x <- numeric(101)
  y <- numeric(101)
  auc <- numeric(101)
  shape <- character(101)

  for (i in 1:101){
    if (i == 1) {xpred[i] <- (i - 1 + 0.001) / 100}
    if (i == 101){xpred[i] <- (i - 1 - 0.001) / 100}
    if (i != 1 && i != 101){xpred[i] <- (i - 1) / 100}
    x[i] <- logit(xpred[i])
    y[i] <- theta_tp + (sigma_tp * sigma_fp * rho)
      / (sigma_fp * sigma_fp) * (x[i] - theta_fp)
  }

  ypred = exp(y) / (1 + exp(y))

  for (i in 1:101){
    if (i == 1){auc[i] <- 0}
    if (i >= 2){auc[i] <- auc[i - 1] + (xpred[i] - xpred[i - 1])
      * (ypred[i] + ypred[i - 1]) / 2}
    shape[i] <- "ROC"
  }

  return(data.frame(xpred = xpred, ypred = ypred, auc = auc,
    shape = shape))
}
```

```
}

# Calculate AUC
fauc <- function(data) {
  sroc1 <- fsroc(theta_tp = data[3, 1],
                 theta_fp = data[4, 1],
                 sigma_tp = data[5, 1],
                 sigma_fp = data[6, 1],
                 rho = data[7, 1])
  return(max(sroc1$auc))
}

fauc0 <- fauc(data = bugs0$summary)
```

付録 C 相対距離, 標準化残差, SROC 曲線の AUC への影

響度

【R プログラム】

```
# Prepare empty data
RDtp <- numeric(n_studies)
RDfp <- numeric(n_studies)
RDave <- numeric(n_studies)
RDsyn <- numeric(n_studies)
RDdor <- numeric(n_studies)
SRtp <- numeric(n_studies)
SRfp <- numeric(n_studies)
SRave <- numeric(n_studies)
SRsyn <- numeric(n_studies)
SRdor <- numeric(n_studies)

# Logit transformation of observed data
ltpr_obs <- logit(TP / n1)
lfpr_obs <- logit(FP / n2)
for(i in 1:n_studies) {
  if (TP[i] == n1[i]) {ltpr_obs[i] <- logit(0.99)}
  if (TP[i] == 0)      {ltpr_obs[i] <- logit(0.01)}
  if (FP[i] == n2[i]) {lfpr_obs[i] <- logit(0.99)}
  if (FP[i] == 0)      {lfpr_obs[i] <- logit(0.01)}
}

# Intra-study variability (Reitsma, 2005)
sigmatp <- 1 / (n1 * (TP / n1) * (FN / n1))
sigmafp <- 1 / (n2 * (FP / n2) * (TN / n2))
for(i in 1:n_studies) {
  if (TP[i] == n1[i]) {sigmatp[i] <- 1 / (n1[i] * 0.99 * 0.01)}
  if (TP[i] == 0)      {sigmatp[i] <- 1 / (n1[i] * 0.01 * 0.99)}
  if (FP[i] == n2[i]) {sigmafp[i] <- 1 / (n2[i] * 0.99 * 0.01)}
  if (FP[i] == 0)      {sigmafp[i] <- 1 / (n2[i] * 0.01 * 0.99)}
```

```

}

# Logit transformation and Intra-study variability of DOR
ldor_obs <- log((TP * TN) / (FP * FN))
sigmador <- (1 / TP) + (1 / FP) + (1 / FN) + (1 / TN)
for(i in 1:n_studies) {
  if (TP[i] == n1[i])
    {ldor_obs[i] <- log((TP[i] * TN[i]) / (FP[i] * 0.1))}
  if (TP[i] == 0)
    {ldor_obs[i] <- log((0.1 * TN[i]) / (FP[i] * FN[i]))}
  if (FP[i] == n2[i])
    {ldor_obs[i] <- log((TP[i] * 0.1) / (FP[i] * FN[i]))}
  if (FP[i] == 0)
    {ldor_obs[i] <- log((TP[i] * TN[i]) / (0.1 * FN[i]))}

  if (TP[i] == n1[i]) {sigmador[i] <- (1 / TP[i]) + (1 / FP[i])
    + (1 / 0.1) + (1 / TN[i])}
  if (TP[i] == 0) {sigmador[i] <- (1 / 0.1) + (1 / FP[i])
    + (1 / FN[i]) + (1 / TN[i])}
  if (FP[i] == n2[i]) {sigmador[i] <- (1 / TP[i]) + (1 / FP[i])
    + (1 / FN[i]) + (1 / 0.1)}
  if (FP[i] == 0) {sigmador[i] <- (1 / TP[i]) + (1 / 0.1)
    + (1 / FN[i]) + (1 / TN[i])}
}

# Relative distance and standardized residuals
for(i in 1:n_studies) {

  # Prepare one-leave-out dataset without study i
  tp_2 <- TP[-i]
  fp_2 <- FP[-i]
  n1_2 <- TP[-i] + FN[-i]
  n2_2 <- TN[-i] + FP[-i]
  n_studies_2 <- length(TP[-i])

```



```

# Making the dataset for OpenBUGS without study i
bugs.data(list("tp" = tp_2, "fp" = fp_2, "n1" = n1_2, "n2" = n2_2,
  "n_studies" = n_studies_2), data.file = "dtadata2.txt" )

# Call OpenBUGS without study i (OpenBUGS プログラムは付録 A を参照)
bugs1 <- bugs(data = "dtadata2.txt",
  inits = list(initv0),
  parameters.to.save = c("tpr_pool", "fpr_pool", "mu0",
    "sigma1", "sigma2", "rho", "ldor_pool", "ltp",
    "lfpr", "ldor"),
  model.file = "Reitsma.txt",
  n.chains = 1,
  n.iter = n_iter,
  n.burnin = n_burnin,
  n.thin = n_thin,
  working.directory = pathname,
  debug = FALSE)

### Relative distance
RDtp[i] <- (logit.inv(bugs0$mean$mu0[1])
  - logit.inv(bugs1$mean$mu0[1])) / logit.inv(bugs0$mean$mu0[1])
RDfp[i] <- (logit.inv(bugs0$mean$mu0[2])
  - logit.inv(bugs1$mean$mu0[2])) / logit.inv(bugs0$mean$mu0[2])
RDave[i] <- (abs(RDtp[i]) + abs(RDfp[i])) / 2
RDSyn[i] <- sqrt((logit.inv(bugs0$mean$mu0[1])
  - logit.inv(bugs1$mean$mu0[1])) ^ 2
  + (logit.inv(bugs0$mean$mu0[2])
  - logit.inv(bugs1$mean$mu0[2])) ^ 2)
  / sqrt(logit.inv(bugs0$mean$mu0[1]) ^ 2
  + logit.inv(bugs0$mean$mu0[2]) ^ 2)
RDdor[i] <- (exp(bugs0$mean$ldor_pool)
  - exp(bugs1$mean$ldor_pool)) / exp(bugs0$mean$ldor_pool)

### Standardized residuals: sensitivity and FPR
se_tp <- sqrt(bugs1$mean$sigma1 ^ 2 + sigmatp[i])

```

```

se_fp <- sqrt(bugs1$mean$sigma2 ^ 2 + sigmafp[i])
se_dor <- sqrt(bugs1$mean$sigma1 ^ 2 + bugs1$mean$sigma2 ^ 2
              - 2 * bugs1$mean$sigma1 * bugs1$mean$sigma2
              * bugs1$mean$rho + sigmador[i])
SRtp[i] <- (ltp_obs[i] - bugs1$mean$mu0[1]) / se_tp
SRfp[i] <- (lfpr_obs[i] - bugs1$mean$mu0[2]) / se_fp
SRave[i] <- (abs(SRtp[i]) + abs(SRfp[i])) / 2
SRdor[i] <- (ldor_obs[i] - bugs1$mean$ldor_pool) / se_dor

### standardized residuals: bi-variate
obs <- rbind(ltp_obs[i], lfpr_obs[i]) # Observed data
est <- rbind(bugs1$mean$mu0[1], bugs1$mean$mu0[2]) # estimation
sig_intra <- rbind(c(sigmatp[i], 0), c(0, sigmafp[i]))
sig_heter <- rbind(c(bugs1$mean$sigma1 ^ 2,
                    bugs1$mean$sigma1 * bugs1$mean$sigma2 * bugs1$mean$rho),
                  c(bugs1$mean$sigma1 * bugs1$mean$sigma2 * bugs1$mean$rho,
                    bugs1$mean$sigma2 ^ 2))
sig <- sig_intra + sig_heter
RSyn[i] <- t(obs - est) %*% solve(sig) %*% (obs - est)

### AUC
fauc1[i] <- fauc(data = bugs1$summary)
aucd[i] <- fauc0 - fauc1[i]
}

```

付録 D ベイズ流 P 値

【R プログラム】

```
# Prepare empty data
Dtp_obs <- numeric(n_studies)
Dfp_obs <- numeric(n_studies)
Dave_obs <- numeric(n_studies)
Dsyn_obs <- numeric(n_studies)
Ddor_obs <- numeric(n_studies)

Dtp_new <- matrix(0:0, nrow = n_studies, ncol = n_mcmc)
Dfp_new <- matrix(0:0, nrow = n_studies, ncol = n_mcmc)
Dave_new <- matrix(0:0, nrow = n_studies, ncol = n_mcmc)
Dsyn_new <- matrix(0:0, nrow = n_studies, ncol = n_mcmc)
Ddor_new <- matrix(0:0, nrow = n_studies, ncol = n_mcmc)

p_tp <- numeric(n_studies)
p_fp <- numeric(n_studies)
p_ave <- numeric(n_studies)
p_syn <- numeric(n_studies)
p_dor <- numeric(n_studies)

# Calculate Bayesian p-value: sensitivity, False positive rate and
these average
for(i in 1:n_studies) {

  # Distance for observed data
  Dtp_obs[i] <- ((ltp_obs[i] - bugs0$mean$ltp[i]) ^ 2)
    / (bugs0$sd$ltp[i] ^ 2)
  Dfp_obs[i] <- ((lfp_obs[i] - bugs0$mean$lfpr[i]) ^ 2)
    / (bugs0$sd$lfpr[i] ^ 2)
  Dave_obs[i] <- (Dtp_obs[i] + Dfp_obs[i]) / 2
  Ddor_obs[i] <- ((ldor_obs[i] - bugs0$mean$ldor[i]) ^ 2)
    / (bugs0$sd$ldor[i] ^ 2)

  # Distances for new data from posterior distribution
  Dtp_new[i, ] <- ((bugs0$sims.list$ltp[, i]
```

```

- bugs0$mean$ltpr[i]) ^ 2) / (bugs0$sd$ltpr[i] ^ 2)
Dfp_new[i, ] <- ((bugs0$sims.list$lfpr[, i]
- bugs0$mean$lfpr[i]) ^ 2) / (bugs0$sd$lfpr[i] ^ 2)
Dave_new[i, ] <- (Dtp_new[i, ] + Dfp_new[i, ]) / 2
Ddor_new[i, ] <- ((bugs0$sims.list$ldor[, i]
- bugs0$mean$ldor[i]) ^ 2) / (bugs0$sd$ldor[i] ^ 2)

# Compare these distances and calculate p-value
p_tp[i] <- sum(Dtp_obs[i] < Dtp_new[i, ]) / n_mcmc
p_fp[i] <- sum(Dfp_obs[i] < Dfp_new[i, ]) / n_mcmc
p_ave[i] <- sum(Dave_obs[i] < Dave_new[i, ]) / n_mcmc
p_dor[i] <- sum(Ddor_obs[i] < Ddor_new[i, ]) / n_mcmc
}

# Calculate Bayesian p-value: bi-variate
for(i in 1:n_studies) {

# Prepare observed data, mean vector, and covariance matrix
obsdata <- rbind(ltpr_obs[i], lfpr_obs[i])
mean <- rbind(bugs0$mean$ltpr[i], bugs0$mean$lfpr[i])
sig <- cov(data.frame(bugs0$sims.list$ltpr[, i],
bugs0$sims.list$lfpr[, i]))

# Distance for observed data
Dsyn_obs[i] <- t(obsdata - mean) %*% solve(sig) %*% (obsdata - mean)

# Distances for new data from posterior distribution
for(k in 1:n_mcmc) {
newdata <- rbind(bugs0$sims.list$ltpr[k, i],
bugs0$sims.list$lfpr[k, i])
Dsyn_new[i, k] <- t(newdata - mean) %*% solve(sig)
%*% (newdata - mean)
}

# Compare these distances and calculate p-value

```

```
p_syn[i] <- sum(Dsyn_obs[i] < Dsyn_new[i, ]) / n_mcmc  
}
```

付録 E 尺度混合正規分布における尺度パラメータの事後

推定値

【R プログラム】

```
# Logit transformation of observed data
fpr <- logit(FP / (TN + FP))
sens <- logit(TP / (TP + FN))

for(i in 1:n_studies) {
  if (TP[i] == n1[i]) {sens[i] <- logit(0.99)}
  if (TP[i] == 0)     {sens[i] <- logit(0.01)}
  if (FP[i] == n2[i]) {fpr[i] <- logit(0.99)}
  if (FP[i] == 0)     {fpr[i] <- logit(0.01)}
}

# Intra-study variability (Reitsma, 2005)
sigmatp <- n1 * (TP / n1) * (FN / n1)
sigmafpr <- n2 * (FP / n2) * (TN / n2)

for(i in 1:n_studies) {
  if (TP[i] == n1[i]) {sigmatp[i] <- 1 / (n1[i] * 0.99 * 0.01)}
  if (TP[i] == 0)     {sigmatp[i] <- 1 / (n1[i] * 0.01 * 0.99)}
  if (FP[i] == n2[i]) {sigmafpr[i] <- 1 / (n2[i] * 0.99 * 0.01)}
  if (FP[i] == 0)     {sigmafpr[i] <- 1 / (n2[i] * 0.01 * 0.99)}
}

### Making the dataset for OpenBUGS
bugs.data(list("sens" = sens, "fpr" = fpr, "sigmatp" = sigmatp,
              "sigmafpr" = sigmafpr, "n_studies" = n_studies),
          data.file = "dtadata.txt")

initv0 <- list(mu0 = c(0, 0), sigma1 = 1, sigma2 = 1, rho = 0)
```

【R プログラム：正規誤差モデル $\lambda_{Ai} = \lambda_{Bi} = 1$ 】

```
bugs1 <- bugs(data = "dtadata.txt",
              inits = list(initv0),
              parameters.to.save = c("mu0", "sigma1", "sigma2", "rho",
                                     "theta"),
              model.file = "Reitsma_normal1.txt",
              n.chains = 1,
              n.iter = n_iter,
              n.burnin = n_burnin,
              n.thin = n_thin,
              working.directory = pathname,
              debug = FALSE)
```

【R プログラム：二重指数分布誤差モデル $\lambda_{Ai}, \lambda_{Bi} \sim EXP(2)$ 】

```
bugs2 <- bugs(data = "dtadata.txt",
              inits = list(initv0),
              parameters.to.save = c("mu0", "sigma1", "sigma2", "rho",
                                     "theta", "lambda1", "lambda2"),
              model.file = "Reitsma_normal2.txt",
              n.chains = 1,
              n.iter = n_iter,
              n.burnin = n_burnin,
              n.thin = n_thin,
              working.directory = pathname,
              debug = FALSE)
```

【R プログラム：t分布誤差モデル $\lambda_{Ai}, \lambda_{Bi} \sim IG(1, 1)$ 】

```
bugs3 <- bugs(data = "dtadata.txt",
              inits = list(initv0),
              parameters.to.save = c("mu0", "sigma1", "sigma2", "rho",
                                     "theta", "lambda1", "lambda2"),
              model.file = "Reitsma_normal3.txt",
              n.chains = 1,
              n.iter = n_iter,
              n.burnin = n_burnin,
```

```
n.thin = n_thin,  
working.directory = pathname,  
debug = FALSE)
```

【R プログラム : CSP モデル $\lambda_i = \lambda_{Ai} = \lambda_{Bi}$, $\lambda_i \sim EXP(2)$ 】

```
bugs4 <- bugs(data = "dtadata.txt",  
             inits = list(initv0),  
             parameters.to.save = c("mu0", "sigma1", "sigma2", "rho",  
                                     "theta", "lambda"),  
             model.file = "Reitsma_normal14.txt",  
             n.chains = 1,  
             n.iter = n_iter,  
             n.burnin = n_burnin,  
             n.thin = n_thin,  
             working.directory = pathname,  
             debug = FALSE)
```

【R プログラム : CSP モデル $\lambda_i = \lambda_{Ai} = \lambda_{Bi}$, $\lambda_i \sim IG(1,1)$ 】

```
bugs5 <- bugs(data = "dtadata.txt",  
             inits = list(initv0),  
             parameters.to.save = c("mu0", "sigma1", "sigma2", "rho",  
                                     "theta", "lambda"),  
             model.file = "Reitsma_normal15.txt",  
             n.chains = 1,  
             n.iter = n_iter,  
             n.burnin = n_burnin,  
             n.thin = n_thin,  
             working.directory = pathname,  
             debug = FALSE)
```

【R プログラム : 各モデルでの結果の統合】

```
SMNp_DE1 <- numeric(n_studies)  
SMNp_DE2 <- numeric(n_studies)  
SMNp_DE3 <- numeric(n_studies)  
SMNp_IG1 <- numeric(n_studies)
```



```

SMNp_IG2 <- numeric(n_studies)
SMNp_IG3 <- numeric(n_studies)

for(j in 1:n_studies) {
  SMNp_DE1[j] <- sum(bugs2$sims.list$lambda1[, j] > 1) / n_mcmc
  SMNp_DE2[j] <- sum(bugs2$sims.list$lambda2[, j] > 1) / n_mcmc
  SMNp_DE3[j] <- sum(bugs4$sims.list$lambda[, j] > 1) / n_mcmc
  SMNp_IG1[j] <- sum(bugs3$sims.list$lambda1[, j] > 1) / n_mcmc
  SMNp_IG2[j] <- sum(bugs3$sims.list$lambda2[, j] > 1) / n_mcmc
  SMNp_IG3[j] <- sum(bugs5$sims.list$lambda[, j] > 1) / n_mcmc
}

result <- round(data.frame(nn, SMNp_IG1, SMNp_DE1, SMNp_IG2, SMNp_DE2,
  SMNp_IG3, SMNp_DE3), 4)

```

【OpenBUGS プログラム：正規誤差モデル $\lambda_{Ai} = \lambda_{Bi} = 1$ 】

```

model {
  for(i in 1:n_studies) {
    # intra-study variability
    sens[i] ~ dnorm(theta[i, 1], sigmatp1[i])
    fpr[i] ~ dnorm(theta[i, 2], sigmafp1[i])

    # heterogeneity
    theta[i, 1:2] ~ dnorm(mu0[1:2], sigma.inv[1:2, 1:2])
  }

  # Priors for means
  mu0[1] ~ dnorm(0, 0.01)
  mu0[2] ~ dnorm(0, 0.01)

  # Priors for the elements of the covariance matrix
  sigma.inv[1:2, 1:2] <- inverse(sigma[1:2, 1:2])
  sigma[1, 1] <- sigma1 * sigma1
  sigma[1, 2] <- sigma1 * sigma2 * rho
  sigma[2, 1] <- sigma1 * sigma2 * rho

```

```

sigma[2, 2] <- sigma2 * sigma2
sigma1 ~ dunif(0.01, 10)
sigma2 ~ dunif(0.01, 10)
rho ~ dunif(-1, 1)

# Pooled summaries
tpr_pool <- exp(mu0[1]) / (1 + exp(mu0[1]))
fpr_pool <- exp(mu0[2]) / (1 + exp(mu0[2]))
}

```

【OpenBUGS プログラム：二重指数分布誤差モデル $\lambda_{Ai}, \lambda_{Bi} \sim EXP(2)$ 】

* intra-study variability 以外の部分は正規誤差モデルと同じである。

```

# intra-study variability
sens[i] ~ dnorm(theta[i, 1], sigmatp1[i])
fpr[i] ~ dnorm(theta[i, 2], sigmafpr1[i])
sigmatp1[i] <- sigmatp[i] / lambda1[i]
sigmafpr1[i] <- sigmafpr[i] / lambda2[i]
lambda1[i] ~ dexp(0.5)
lambda2[i] ~ dexp(0.5)

```

【OpenBUGS プログラム：t 分布誤差モデル $\lambda_{Ai}, \lambda_{Bi} \sim IG(1, 1)$ 】

* intra-study variability 以外の部分は正規誤差モデルと同じである。

```

# intra-study variability
sens[i] ~ dnorm(theta[i, 1], sigmatp1[i])
fpr[i] ~ dnorm(theta[i, 2], sigmafpr1[i])
sigmatp1[i] <- sigmatp[i] / lambda1[i]
sigmafpr1[i] <- sigmafpr[i] / lambda2[i]
lambda1[i] <- 1 / lambda.inv1[i]
lambda2[i] <- 1 / lambda.inv2[i]
lambda.inv1[i] ~ dgamma(1, 1)
lambda.inv2[i] ~ dgamma(1, 1)

```

【OpenBUGS プログラム：CSP モデル $\lambda_i = \lambda_{Ai} = \lambda_{Bi}$, $\lambda_i \sim EXP(2)$ 】

* intra-study variability 以外の部分は正規誤差モデルと同じである。

```

# intra-study variability

```

```

sens[i] ~ dnorm(theta[i, 1], sigmatp1[i])
fpr[i] ~ dnorm(theta[i, 2], sigmafp1[i])
sigmatp1[i] <- sigmatp[i] / lambda[i]
sigmafp1[i] <- sigmafp[i] / lambda[i]
lambda[i] ~ dexp(0.5)

```

【OpenBUGS プログラム : CSP モデル $\lambda_i = \lambda_{Ai} = \lambda_{Bi}$, $\lambda_i \sim IG(1, 1)$ 】

* intra-study variability 以外の部分は正規誤差モデルと同じである。

```
# intra-study variability
```

```

sens[i] ~ dnorm(theta[i, 1], sigmatp1[i])
fpr[i] ~ dnorm(theta[i, 2], sigmafp1[i])
sigmatp1[i] <- sigmatp[i] / lambda[i]
sigmafp1[i] <- sigmafp[i] / lambda[i]
lambda[i] <- 1 / lambda.inv[i]
lambda.inv[i] ~ dgamma(1, 1)

```

謝辞

博士課程におきまして、主任指導教員としてご指導くださり、常に励ましてくださりました野間久史准教授に、深くお礼申し上げます。

学位審査に関しましては、ご多忙にもかかわらず、博士論文の審査委員をお引き受け下さりました伊庭幸人教授、逸見昌之准教授、大阪大学の服部聡教授に謹んで感謝申し上げます。博士論文の審査の過程で、多くの有益なご指摘を頂きました。また、逸見昌之准教授には、副指導教員もご担当頂きました。重ねて、感謝申し上げます。

本研究の投稿論文に関しましては、京都大学の古川壽亮教授、King's College London および東京大学の山田朋英先生には、共著者として、大変有益なご助言、ご指導を頂きました。篤くお礼申し上げます。

また、私の学生生活を様々な面から支えてくださりました、統計数理研究所の皆様にお礼申し上げます。

最後に、勤務先である大塚製薬株式会社 新薬開発本部の皆様と、私の家族に感謝いたします。

参考文献

1. Deeks, J. J. (2001). Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests, *BMJ*, **323**(7305), 157-162.
2. Leeflang, M. M., Deeks, J. J., Gatsonis, C. and Bossuyt, P. M. (2008). Systematic reviews of diagnostic test accuracy, *Annals of Internal Medicine*, **149**(12), 889-897.
3. Harbord, R. M., Deeks, J. J., Egger, M., Whiting, P. and Sterne, J. A. (2007). A unification of models for meta-analysis of diagnostic accuracy studies, *Biostatistics*, **8**(2), 239-251.
4. Reitsma, J. B., Glas, A. S., Rutjes, A. W., Scholten, R. J., Bossuyt, P. M. and Zwinderman, A. H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews, *Journal of Clinical Epidemiology*, **58**(10), 982-990.
5. Rutter, C. M. and Gatsonis, C. A. (2001). A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations, *Statistics in Medicine*, **20**(19), 2865-2884.
6. van Houwelingen, H. C., Arends, L. R. and Stijnen, T. (2002). Advanced methods in meta-analysis: multivariate approach and meta-regression, *Statistics in Medicine*, **21**(4), 589-624.
7. van Houwelingen, H. C., Zwinderman, K. H. and Stijnen, T. (1993). A bivariate approach to meta-analysis, *Statistics in Medicine*, **12**(24), 2273-2284.
8. Arends, L. R., Hamza, T. H., van Houwelingen, J. C., Heijenbrok-Kal, M. H., Hunink, M. G. and Stijnen, T. (2008). Bivariate random effects meta-analysis of ROC curves, *Medical Decision Making*, **28**(5), 621-638.

9. Verde, P. E. (2010). Meta-analysis of diagnostic test data: a bivariate Bayesian modeling approach, *Statistics in Medicine*, **29**(30), 3088-3102.
10. Verde, P. E. (2018). bamdit: an R package for Bayesian meta-analysis of diagnostic test data, *Journal of Statistical Software*, **86**(10).
11. Dwamena, B. A. (2007). *MIDAS: Stata module for meta-analytical integration of diagnostic test accuracy studies*, Statistical Software Components S456880, Boston College Department of Economics.
12. Harbord, R. M. and Whiting, P. (2009). metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression, *The Stata Journal*, **9**(2), 211-229.
13. 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.
14. Carlin, B. P. and Louis, T. A. (2008). *Bayesian Methods for Data Analysis*, Chapman and Hall/CRC Press, Boca Raton, Florida.
15. Zhang, J., Fu, H. and Carlin, B. P. (2015). Detecting outlying trials in network meta-analysis, *Statistics in Medicine*, **34**(19), 2695-2707.
16. Chu, H. and Cole, S. R. (2006). Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach, *Journal of Clinical Epidemiology*, **59**(12), 1331-1333.
17. 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.
18. Korevaar, D. A., Westerhof, G. A., Wang, J., Cohen, J. F., Spijker, R., Sterk, P. J., Bel, E. H. and Bossuyt, P. M. M. (2015). Diagnostic accuracy of minimally

- invasive markers for detection of airway eosinophilia in asthma: a systematic review and meta-analysis, *The Lancet Respiratory Medicine*, **3**(4), 290-300.
19. Matsushima, Y., Noma, H., Yamada, T. and Furukawa, T. A. (2020). Influence diagnostics and outlier detection for meta-analysis of diagnostic test accuracy, *Research Synthesis Methods*, **11**(2), 237-247.
 20. Lunn, D., Jackson, C., Best, N., Thomas, A. and Spiegelhalter, D. (2012). *The BUGS Book: A Practical Introduction to Bayesian Analysis*, Chapman and Hall/CRC, Boca Raton, Florida.
 21. Lunn, D., Spiegelhalter, D., Thomas, A. and Best, N. (2009). The BUGS project: Evolution, critique and future directions, *Statistics in Medicine*, **28**(25), 3049-3067.
 22. Glas, A. S., Lijmer, J. G., Prins, M. H., Bonsel, G. J. and Bossuyt, P. M. (2003). The diagnostic odds ratio: a single indicator of test performance, *Journal of Clinical Epidemiology*, **56**(11), 1129-1135.
 23. Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2013). *Bayesian Data Analysis*, Chapman and Hall/CRC Press, Boca Raton, Florida.
 24. Andrews, D. F. and Mallows, C. L. (1974). Scale mixtures of normal distributions, *Journal of the Royal Statistical Society Series B (Methodological)*, **36**(1), 99-102.
 25. R Core Team. (2018). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
 26. Viechtbauer, W. and Cheung, M. W. (2010). Outlier and influence diagnostics for meta-analysis, *Research Synthesis Methods*, **1**(2), 112-125.
 27. Noma, H., Matsushima, Y. and Ishii, R. (2020). Confidence interval for the AUC of SROC curve and some related methods using bootstrap for meta-analysis of

diagnostic accuracy studies, arXiv: 2004.04339.