数理解析研究所講究録
第1804巻 2012年 30-39

複数の評価者による対応のあるクラスターデータの割合の差の検定

佐伯 浩之[1,3]、丹後 俊郎[2]、汪 金芳[1]

[1] 千葉大学大学院 理学研究科
[2] 医学統計学研究センター
[3] 富士フイルムRIファーマ株式会社

1 はじめに

二つの診断法を比較する臨床試験において、癌の骨転移巣のように一人の被験者に存在する複数の病変を解析対象とするような、クラスターデータの構造のもとでの対応の割合を比較する際には、データのクラスター内相関を適切に考慮する必要がある。このようなクラスターデータで構成された割合の差の検定手法としては、Obuchowski [1]、Durkalski ら [2]、Nam and Kwon [3] 及び Jin and Lu [4] の方法が適用可能である。ただし、これらの方法は、二つの診断法が1名の評価者によって評価された場合にのみ利用できる方法である。一般に、診断法を比較する臨床試験では、診断法以外の情報の混入によるバイアスの除去、及び評価の信頼性を確認することを目的として、臨床試験の実施施設とは独立した複数の評価者による結果を利用することが多い (Lehr and Kashanian [5])。複数の評価者による結果については、合議や多数決を用いて単独の評価者による結果として取り扱うことで、上記の検定手法を適用可能である。しかしながら、合議は非独立の評価であるためにバイアスを生じる可能性があり、多数決は複数の評価者からの結果のバラツキを考慮できないことから、これらの方法は主要な評価に対して推奨されていない (FDA [6]、Obuchowski and Lieber [7]、CHMP [8])。従って、複数の評価者から得た結果の全てを解析に利用することが望まれる。Saeki and Tango [9] は、複数の評価者による対応のある割合の差に対する非劣性検定を報告しているが、本手法をクラスターデータに利用することは適切ではない。そこで本研究では、複数の評価者から得た全ての結果を利用する、対応のあるクラスターデータの割合の差の検定法を提案する。

2 統計モデル

同一の被験者 j に独立に実施された従来法 (Standard method) と新規法 (New method) の画像診断を、n_j 個の病変に対して K 名の評価者が独立に評価した結果に基づいて比較する状況を考える。各評価者は、陽性 (+) 又は陰性 (−) の判定を行うこととする。このような状況において得られるクラスターデータを、本研究で対象とするデータ構造として設定する。このデータセットの構造を表1に示す。

最もシンプルなケースとして評価者2名の状況では、クラスター j から得られたデータを4×4分割表として示すことができる (表2)。例えば、r_{1101j} は、「評価者1が新規法を陽性、評価者2が新規法を陽性、評価者1が従来法を陰性、評価者2が従来法を陽性」と判定する確率で、y_{1101j} はその観測度数を表している。

ここで、$\pi_{Nj}^{(k)}(\pi_{Sj}^{(k)})$ を、クラスター j における新規法 (又は従来法) での評価者 k の陽性確率とおく。次に、π_{Nj} と π_{Sj} の各々をクラスター j における新規法及び従来法での陽性確率と定義し、さらにこれら確率を以下のように定式化する。

$$\pi_{Nj} = \omega^{(1)}\pi_{Nj}^{(1)} + \omega^{(2)}\pi_{Nj}^{(2)} \tag{1}$$

$$\pi_{Sj} = \omega^{(1)}\pi_{Sj}^{(1)} + \omega^{(2)}\pi_{Sj}^{(2)} \tag{2}$$

表1　画像診断の比較におけるクラスターデータのデータセット構造

Subject ID	Method	lesion ID	Rater 1	Rater 2	.....	Rater K
1	Standard	1	+	+	.....	+
1	Standard	2	−	+	.....	+
1	Standard	:	+	+	.....	+
1	Standard	n_1	+	+	.....	−
1	New	1	−	−	.....	+
1	New	2	+	+	.....	+
1	New	:	+	−	.....	+
1	New	n_1	+	+	.....	+
2	Standard	1	+	+	.....	+
2	Standard	:	−	+	.....	+
2	Standard	n_2	+	+	.....	+
2	New	1	+	+	.....	+
2	New	:	−	+	.....	+
2	New	n_2	+	+	.....	+
:	:	:	:	:	.....	:
j	Standard	1	−	−	.....	−
j	Standard	:	−	+	.....	−
j	Standard	n_j	−	−	.....	−
j	New	1	+	+	.....	+
j	New	:	−	−	.....	−
j	New	n_j	+	+	.....	+

ここで $\omega^{(1)}$ と $\omega^{(2)}$ $(\omega^{(1)}+\omega^{(2)}=1)$ を、$\pi_{Nj}^{(1)}$ $(\pi_{Sj}^{(1)})$ 及び $\pi_{Nj}^{(2)}$ $(\pi_{Sj}^{(2)})$ の π_{Nj} (π_{Sj}) に寄与する重みとする。もし、各評価者の評価レベルが等しければ、この重みは評価者間で等しいとみなすことができるので、これら確率は $K=2$ の状況において以下のように示すことが可能となる。

$$\pi_{Nj}=\frac{\pi_{Nj}^{(1)}+\pi_{Nj}^{(2)}}{2}=r_{11\cdot\cdot j}+\frac{r_{10\cdot\cdot j}+r_{01\cdot\cdot j}}{2} \tag{3}$$

表2　クラスター j (j=1, 2,..., m) において、二つの方法を二名の評価者が二値の判定を行う場合の 4×4 分割表

		Standard Method				
	Judgment of (Rater 1, Rater 2)	(+,+)	(+,−)	(−,+)	(−,−)	Total
New Method	(+,+)	r_{1111j} (y_{1111j})	r_{1110j} (y_{1110j})	r_{1101j} (y_{1101j})	r_{1100j} (y_{1100j})	$r_{11\cdot\cdot j}$ $(y_{11\cdot\cdot j})$
	(+, −)	r_{1011j} (y_{1011j})	r_{1010j} (y_{1010j})	r_{1001j} (y_{1001j})	r_{1000j} (y_{1000j})	$r_{10\cdot\cdot j}$ $(y_{10\cdot\cdot j})$
	(−, +)	r_{0111j} (y_{0111j})	r_{0110j} (y_{0110j})	r_{0101j} (y_{0101j})	r_{0100j} (y_{0100j})	$r_{01\cdot\cdot j}$ $(y_{01\cdot\cdot j})$
	(−, −)	r_{0011j} (y_{0011j})	r_{0010j} (y_{0010j})	r_{0001j} (y_{0001j})	r_{0000j} (y_{0000j})	$r_{00\cdot\cdot j}$ $(y_{00\cdot\cdot j})$
	Total	$r_{\cdot\cdot 11j}$ $(y_{\cdot\cdot 11j})$	$r_{\cdot\cdot 10j}$ $(y_{\cdot\cdot 10j})$	$r_{\cdot\cdot 01j}$ $(y_{\cdot\cdot 01j})$	$r_{\cdot\cdot 00j}$ $(y_{\cdot\cdot 00j})$	1 (n_j)

$$\pi_{Sj} = \frac{\pi_{Sj}^{(1)} + \pi_{Sj}^{(2)}}{2} = r_{\cdot\cdot 11j} + \frac{r_{\cdot\cdot 10j} + r_{\cdot\cdot 01j}}{2} \tag{4}$$

さらに、式 (3) 及び式 (4) に基づき、4×4 分割表は、3×3 分割表として置き換えることができる (表 3)。表 3 における p_{lmj} (x_{lmj}) を、クラスター j において新規法では l 名の評価者が陽性と判定し、従来法では m 名の評価者が陽性と判定した確率 (観測度数) と定義する。

表 3　クラスター j (j=1, 2,..., m) において、二つの方法を二名の評価者が二値の判定を行う場合の 3×3 分割表

		Standard Method			
	Judgment of (Rater 1, Rater 2)	(+, +)	(+, −) or (−,+)	(−, −)	Total
New Method	(+, +)	p_{22j} (x_{22j})	p_{21j} (x_{21j})	p_{20j} (x_{20j})	$p_{2\cdot j}$ $(x_{2\cdot j})$
	(+, −) or (−, +)	p_{12j} (x_{12j})	p_{11j} (x_{11j})	p_{10j} (x_{10j})	$p_{1\cdot j}$ $(x_{1\cdot j})$
	(−, −)	p_{02j} (x_{02j})	p_{01j} (x_{01j})	p_{00j} (x_{00j})	$p_{0\cdot j}$ $(x_{0\cdot j})$
	Total	$p_{\cdot 2j}$ $(x_{\cdot 2j})$	$p_{\cdot 1j}$ $(x_{\cdot 1j})$	$p_{\cdot 0j}$ $(x_{\cdot 0j})$	1 (n_j)

従って、表 3 の様式に基づき π_{Nj} 及び π_{Sj} はそれぞれ以下のように示される。

$$\begin{aligned}\pi_{Nj} &= p_{2\cdot j} + \frac{1}{2}p_{1\cdot j} \\ &= p_{20j} + (p_{21j} + \frac{1}{2}p_{10j}) + (p_{22j} + \frac{1}{2}p_{11j}) + \frac{1}{2}p_{12j}\end{aligned} \tag{5}$$

$$\begin{aligned}\pi_{Sj} &= p_{\cdot 2j} + \frac{1}{2}p_{\cdot 1j} \\ &= p_{02j} + (p_{12j} + \frac{1}{2}p_{01j}) + (p_{22j} + \frac{1}{2}p_{11j}) + \frac{1}{2}p_{21j}\end{aligned} \tag{6}$$

次に、二種類の診断法における陽性確率の差 λ_j と、その推定値 $\tilde{\lambda}_j$ を以下のように定義する。

$$\begin{aligned}\lambda_j &= \pi_{Nj} - \pi_{Sj} \\ &= p_{20j} + \frac{1}{2}(p_{21j} + p_{10j}) - p_{02j} - \frac{1}{2}(p_{12j} + p_{01j})\end{aligned} \tag{7}$$

$$\tilde{\lambda}_j = \frac{1}{n}\left\{x_{20j} + \frac{1}{2}(x_{21j} + x_{10j}) - x_{02j} - \frac{1}{2}(x_{12j} + x_{01j})\right\} \tag{8}$$

ここで、x_{20j} は従来法に比べて新規法で陽性と判定した評価者の人数が 2 名多い頻度、$(x_{21j} + x_{10j})$ は従来法に比べて新規法で陽性と判定した評価者の人数が 1 名多い頻度である。同様に、x_{02j} は新規法に比べて従来法で陽性と判定した評価者の人数が 2 名多い頻度、$(x_{12j} + x_{01j})$ は新規法に比べて従来法で陽性と判定した評価者の人数が 1 名多い頻度である。これらの特徴に基づいて評価者の人数を K 名に一般化したとき、対応のあるカテゴリカルデータは $(K+1) \times (K+1)$ 分割表として示すことが可能となる。さらに、先の標記様式については、n_{Nkj} 及び q_{Nkj} を「従来法に比べて新規法で陽性と判定した評価者の人数が k 名多い頻度及びそれに対応する確率」、n_{Skj} 及び q_{Skj} を「新規法に比べて従来法で陽性と判定した評価者の人数が k 名多い頻度及びそれに対応する確率」として簡略化し、以下のように定式化できる。

$$n_{Nkj} = \sum_{l-m=k} x_{lmj}$$

$$q_{Nkj} = \sum_{l-m=k} p_{lmj}$$

$$n_{Skj} = \sum_{l-m=-k} x_{lmj}$$

$$q_{Skj} = \sum_{l-m=-k} p_{lmj}$$

以上より、クラスター j (j=1, 2,..., m) における π_{Nj} 及び π_{Sj} は次のようになる。

$$\pi_{Nj} = \frac{1}{K}\sum_{k=1}^{K}\pi_{Nj}^{(k)}, \quad \pi_{Sj} = \frac{1}{K}\sum_{k=1}^{K}\pi_{Sj}^{(k)} \tag{9}$$

陽性確率の差 λ_j とその推定値 $\tilde{\lambda}_j$ は、

$$\lambda_j = \pi_{Nj} - \pi_{Sj} = \frac{1}{K}\sum_{k=1}^{K} kp_{k\cdot j} - \frac{1}{K}\sum_{k=1}^{K} kp_{\cdot kj}$$

$$= \frac{1}{K}\sum_{k=1}^{K} k(q_{Nkj} - q_{Skj}) \tag{10}$$

$$\tilde{\lambda}_j = \frac{1}{K}\sum_{k=1}^{K} k(\tilde{q}_{Nkj} - \tilde{q}_{Skj})$$

$$= \frac{1}{n_j K}\sum_{k=1}^{K} k(n_{Nkj} - n_{Skj}) \tag{11}$$

として示される。ここで、$\tilde{q}_{Nkj}$ 及び $\tilde{q}_{Skj}$ は q_{Nkj} 及び q_{Skj} の推定値、n_j はクラスター j のユニット数である。最終的に、クラスターを統合した陽性確率 λ とその推定値 $\tilde{\lambda}$ は、

$$\lambda = \pi_N - \pi_S = \frac{1}{m}\sum_{j=1}^{m}\left\{\frac{1}{K}\sum_{k=1}^{K} k(q_{Nkj} - q_{Skj})\right\} \tag{12}$$

$$\tilde{\lambda} = \frac{1}{m}\sum_{j=1}^{m}\left\{\frac{1}{K}\sum_{k=1}^{K} k(\tilde{q}_{Nkj} - \tilde{q}_{Skj})\right\}$$

$$= \frac{1}{m}\sum_{j=1}^{m}\left\{\frac{1}{n_j K}\sum_{k=1}^{K} k(n_{Nkj} - n_{Skj})\right\} \tag{13}$$

として示される。

3 割合の差の検定

本章では、クラスター j における π_{Nj} 及び π_{Sj} に基づき、m 名の被験者（クラスター）から得られたデータに対して、K 名の評価者による二つの画像診断の比較を行うための検定統計量を誘導する。ここで、帰無仮説と対立仮説を

$$H_0: \lambda_j = \pi_{Nj} - \pi_{Sj} = 0, \quad H_1: \lambda_j = \pi_{Nj} - \pi_{Sj} \neq 0$$

として考える。

クラスターを統合した陽性確率の差の推定値 $\tilde{\lambda}$ の分散は、式 (13) から多項分布に基づき、以下のように求めることができる。

$$\mathrm{Var}(\tilde{\lambda}) = \frac{1}{m^2}\mathrm{Var}\left[\sum_{j=1}^{m}\left\{\frac{1}{K}\sum_{k=1}^{K} k(\tilde{q}_{Nkj} - \tilde{q}_{Skj})\right\}\right]$$

$$= \frac{1}{m^2}\sum_{j=1}^{m}\left\{\frac{1}{n_j K^2}(A - B + C)\right\}$$

$$= \frac{1}{m^2}\sum_{j=1}^{m}\left[\frac{1}{n_j K^2}\sum_{k=1}^{K} k^2(\tilde{q}_{Nkj} + \tilde{q}_{Skj}) - \frac{1}{n_j}\left\{\frac{1}{K}\sum_{k=1}^{K} k(\tilde{q}_{Nkj} - \tilde{q}_{Skj})\right\}^2\right] \tag{14}$$

ただし、

$$A=\sum_{k=1}^{K}k^2(\tilde{q}_{Nkj}+\tilde{q}_{Skj}-\tilde{q}_{Nkj}^2-\tilde{q}_{Skj}^2)$$

$$B=\sum_{s,t\in K,s<t}2st(\tilde{q}_{Nsj}\tilde{q}_{Ntj}+\tilde{q}_{Ssj}\tilde{q}_{Stj})$$

$$C=\sum_{u,u'\in K}2uu'\tilde{q}_{Nuj}\tilde{q}_{Su'j}$$

である。ここで、帰無仮説 $H_0:\lambda_j=0$ に基づき、

$$\tilde{\lambda}_j=\frac{1}{K}\sum_{k=1}^{K}k(\tilde{q}_{Nkj}-\tilde{q}_{Skj})=0 \tag{15}$$

とみなすと、式 (14) の第二項に式 (15) を代入することで、帰無仮説のもとでの $\tilde{\lambda}$ の分散は以下のように示すことができる。

$$\mathrm{Var}(\tilde{\lambda})_{H_0}=\frac{1}{m^2}\sum_{j=1}^{m}\frac{1}{n_j^2}\left\{\frac{1}{K^2}\sum_{k=1}^{K}k^2(n_{Nkj}+n_{Skj})\right\} \tag{16}$$

従って、Wald-type 検定統計量は式 (13) 及び式 (16) により、以下のように求められる。

$$\begin{aligned}Z_W&=\frac{\tilde{\lambda}}{\sqrt{\mathrm{Var}(\tilde{\lambda})_{H_0}}}=\frac{\frac{1}{m}\sum_{j=1}^{m}\frac{1}{n_jK}\left\{\sum_{k=1}^{K}k(n_{Nkj}-n_{Skj})\right\}}{\sqrt{\frac{1}{m^2}\sum_{j=1}^{m}\frac{1}{n_j^2}\left\{\frac{1}{K^2}\sum_{k=1}^{K}k^2(n_{Nkj}+n_{Skj})\right\}}}\\&=\frac{\sum_{j=1}^{m}\frac{1}{n_j}\left\{\sum_{k=1}^{K}k(n_{Nkj}-n_{Skj})\right\}}{\sqrt{\sum_{j=1}^{m}\frac{1}{n_j^2}\left\{\sum_{k=1}^{K}k^2(n_{Nkj}+n_{Skj})\right\}}}\end{aligned} \tag{17}$$

適当な正則条件のもとで、Z_W は近似的に標準正規分布に従う。

4 数値実験

本研究で誘導した Wald-type 検定統計量 Z_W について、少ない標本数での検定のサイズと検出力を検討するために、有意水準を両側 5%、評価者 K を 2 又は 3、クラスター数 m を 25、50 又は 100、ユニット数 n_j を ≤ 5、≤ 10 又は ≤ 15 の範囲内でランダムに設定した場合と 5、10 又は 15 で固定した場合における、繰り返し回数 10,000 回でのモンテカルロ・シミュレーションを実施した。シミュレーションデータは、パラメータ値 $(q_{N3j},q_{N2j},q_{N1j},q_{S3j},q_{S2j},q_{S1j})$ に対して一般的な状況を考慮したうえで、多項分布に基づき作成した。ここで、パラメータ値は分割表における非対角セルにおける確率の条件を示しており、このパラメータ値が小さいほどクラスター内の相関が高くなる。シミュレーションの結果を表 4、5、6 及び 7 に示す。

Wald-type 検定統計量 Z_W に基づく検定のサイズは、評価者数、クラスター数、ユニット数及びクラスター内におけるデータの相関関係を変化させても、名目上の水準である 5% を概ね保っていた。一方、クラスター数が多い場合、ユニット数が多い場合、及びクラスター内における相関が高い場合には検出力が高くなる傾向が認められた。

5 考察

本研究では、二つの診断法を比較する臨床試験において、試験実施施設とは独立した複数の評価者から得られたクラスターデータを総合的に利用できる、対応のあるクラスターデータの割合の差の検定法を誘導した。本研究で誘導した Wald-type 検定統計量の検定のサイズは、クラスター数、ユニット数及びクラスター内相関によらず、名目上の水準を保っていた。一方、検出力はクラスター数、ユニット数及びクラスター内相関の状況に依存して変動した。

表4　評価者2名、ユニット数 ≤ 5、≤ 10、≤ 15 における有意水準 両側5% でのサイズと検出力

					Power (%)	
n_j	m	$q_{N1j} = q_{S1j}$	$q_{N2j} = q_{S2j}$	Size (%)	$q_{N2j} + 0.1$	$q_{N2j} + 0.05$
≤ 5	100	0.01	0.05	4.9	98.2	62.5
		0.05	0.05	4.3	97.2	57.4
		0.05	0.1	4.8	88.2	39.2
		0.1	0.1	4.6	86.1	36.9
	50	0.01	0.05	5.2	82.0	35.7
		0.05	0.05	4.7	76.8	31.5
		0.05	0.1	5.3	60.4	22.0
		0.1	0.1	4.8	56.6	20.4
	25	0.01	0.05	4.8	52.0	19.7
		0.05	0.05	4.7	46.6	17.1
		0.05	0.1	5.1	34.5	12.9
		0.1	0.1	4.6	32.9	12.1
≤ 10	100	0.01	0.05	5.0	99.8	79.9
		0.05	0.05	4.8	99.6	75.1
		0.05	0.1	4.8	97.0	54.8
		0.1	0.1	4.7	95.7	50.9
	50	0.01	0.05	4.8	93.9	51.4
		0.05	0.05	5.2	91.4	45.0
		0.05	0.1	4.9	78.0	31.4
		0.1	0.1	4.8	74.7	28.8
	25	0.01	0.05	4.7	70.2	29.6
		0.05	0.05	4.9	64.8	26.0
		0.05	0.1	5.0	49.5	18.1
		0.1	0.1	4.8	46.4	16.8
≤ 15	100	0.01	0.05	4.9	100	88.7
		0.05	0.05	5.0	99.9	84.9
		0.05	0.1	5.0	99.1	65.9
		0.1	0.1	4.8	98.8	61.8
	50	0.01	0.05	4.9	97.8	61.5
		0.05	0.05	5.2	96.4	56.5
		0.05	0.1	4.9	87.4	39.1
		0.1	0.1	4.7	84.9	36.6
	25	0.01	0.05	4.6	80.3	36.8
		0.05	0.05	4.6	76.9	32.9
		0.05	0.1	4.8	60.2	22.7
		0.1	0.1	5.0	57.1	20.6

表5　評価者2名、ユニット数5、10、15における有意水準 両側5% でのサイズと検出力

					Power (%)	
n_j	m	$q_{N1j} = q_{S1j}$	$q_{N2j} = q_{S2j}$	Size (%)	$q_{N2j} + 0.1$	$q_{N2j} + 0.05$
5	100	0.01	0.05	5.0	99.9	81.7
		0.05	0.05	5.2	99.8	77.2
		0.05	0.1	5.2	97.7	56.9
		0.1	0.1	4.8	96.9	53.6
	50	0.01	0.05	5.1	95.1	52.2
		0.05	0.05	5.1	93.0	48.1
		0.05	0.1	5.0	80.0	32.9
		0.1	0.1	5.0	77.4	29.6
	25	0.01	0.05	5.0	72.1	29.8
		0.05	0.05	4.9	67.5	26.8
		0.05	0.1	5.0	50.6	18.6
		0.1	0.1	5.2	47.9	17.4
10	100	0.01	0.05	4.8	100	98.3
		0.05	0.05	4.9	100	97.2
		0.05	0.1	5.1	100	86.3
		0.1	0.1	4.9	100	82.2
	50	0.01	0.05	4.7	100	81.4
		0.05	0.05	4.7	99.8	77.0
		0.05	0.1	4.7	98.1	57.2
		0.1	0.1	5.1	97.0	54.0
	25	0.01	0.05	5.0	95.3	51.4
		0.05	0.05	4.7	93.1	47.0
		0.05	0.1	4.8	79.9	32.4
		0.1	0.1	4.8	77.2	30.7
15	100	0.01	0.05	4.7	100	99.9
		0.05	0.05	5.1	100	99.8
		0.05	0.1	4.9	100	96.3
		0.1	0.1	5.0	100	94.8
	50	0.01	0.05	5.0	100	94.0
		0.05	0.05	4.9	100	91.3
		0.05	0.1	4.8	99.8	74.3
		0.1	0.1	4.5	99.8	70.6
	25	0.01	0.05	4.8	99.3	69.7
		0.05	0.05	4.8	98.8	63.8
		0.05	0.1	4.6	92.9	45.0
		0.1	0.1	4.5	91.2	41.5

表6　評価者3名、ユニット数 ≤ 5、≤ 10、≤ 15 における有意水準 両側5% でのサイズと検出力

n_j	m	$q_{N1j}=q_{S1j}$	$q_{N2j}=q_{S2j}$	$q_{N3j}=q_{S3j}$	Size (%)	Power (%)	
						$q_{N3j}+0.1$	$q_{N3j}+0.05$
≤ 5	100	0.01	0.01	0.05	5.1	97.9	61.3
		0.01	0.05	0.05	5.1	95.3	52.1
		0.05	0.05	0.05	4.8	94.7	49.8
		0.05	0.05	0.1	5.0	84.7	35.4
		0.05	0.1	0.1	5.2	79.6	32.2
	50	0.01	0.01	0.05	4.9	79.2	34.3
		0.01	0.05	0.05	4.8	72.7	28.4
		0.05	0.05	0.05	4.7	70.7	28.0
		0.05	0.05	0.1	5.4	55.7	19.9
		0.05	0.1	0.1	5.0	50.3	17.7
	25	0.01	0.01	0.05	5.0	50.2	19.1
		0.01	0.05	0.05	4.9	43.4	15.9
		0.05	0.05	0.05	4.5	42.5	16.4
		0.05	0.05	0.1	5.3	31.3	13.0
		0.05	0.1	0.1	4.8	29.2	11.7
≤ 10	100	0.01	0.01	0.05	4.6	99.9	77.8
		0.01	0.05	0.05	4.9	99.4	68.8
		0.05	0.05	0.05	5.0	99.4	66.9
		0.05	0.05	0.1	5.2	95.8	51.5
		0.05	0.1	0.1	5.2	93.0	44.9
	50	0.01	0.01	0.05	4.8	93.2	48.3
		0.01	0.05	0.05	5.0	88.4	41.6
		0.05	0.05	0.05	4.8	87.2	39.5
		0.05	0.05	0.1	4.9	74.7	29.7
		0.05	0.1	0.1	5.0	68.0	25.6
	25	0.01	0.01	0.05	4.3	69.2	26.8
		0.01	0.05	0.05	4.6	60.7	23.4
		0.05	0.05	0.05	5.0	59.7	22.6
		0.05	0.05	0.1	5.1	45.5	17.3
		0.05	0.1	0.1	5.1	41.1	14.8
≤ 15	100	0.01	0.01	0.05	4.4	100	87.9
		0.01	0.05	0.05	4.7	99.8	79.8
		0.05	0.05	0.05	4.9	99.9	78.1
		0.05	0.05	0.1	5.0	98.6	61.5
		0.05	0.1	0.1	5.2	97.7	55.6
	50	0.01	0.01	0.05	4.5	97.4	61.0
		0.01	0.05	0.05	5.0	94.7	51.0
		0.05	0.05	0.05	4.7	93.8	49.1
		0.05	0.05	0.1	4.8	84.1	36.2
		0.05	0.1	0.1	5.2	79.7	31.6
	25	0.01	0.01	0.05	4.6	80.1	35.0
		0.01	0.05	0.05	4.9	72.8	29.4
		0.05	0.05	0.05	4.3	70.7	28.5
		0.05	0.05	0.1	4.9	55.6	20.1
		0.05	0.1	0.1	4.6	51.1	18.8

表7　評価者3名、ユニット数5、10、15における有意水準 両側5% でのサイズと検出力

						Power (%)	
n_j	m	$q_{N1j} = q_{S1j}$	$q_{N2j} = q_{S2j}$	$q_{N3j} = q_{S3j}$	Size (%)	$q_{N3j} + 0.1$	$q_{N3j} + 0.05$
5	100	0.01	0.01	0.05	5.1	99.9	80.1
		0.01	0.05	0.05	4.8	99.7	72.7
		0.05	0.05	0.05	5.2	99.5	70.3
		0.05	0.05	0.1	4.5	96.9	52.7
		0.05	0.1	0.1	5.0	94.6	48.6
	50	0.01	0.01	0.05	4.8	94.5	51.4
		0.01	0.05	0.05	4.6	90.7	43.5
		0.05	0.05	0.05	5.0	89.1	42.3
		0.05	0.05	0.1	4.9	77.3	30.4
		0.05	0.1	0.1	5.0	70.6	27.6
	25	0.01	0.01	0.05	4.6	71.0	28.6
		0.01	0.05	0.05	5.1	62.8	23.8
		0.05	0.05	0.05	4.8	61.7	23.3
		0.05	0.05	0.1	4.7	47.3	17.4
		0.05	0.1	0.1	4.8	42.8	15.6
10	100	0.01	0.01	0.05	4.8	100	97.9
		0.01	0.05	0.05	4.8	100	95.3
		0.05	0.05	0.05	4.8	100	94.3
		0.05	0.05	0.1	4.6	100	81.9
		0.05	0.1	0.1	4.9	99.9	75.9
	50	0.01	0.01	0.05	4.6	100	80.7
		0.01	0.05	0.05	4.7	99.7	72.2
		0.05	0.05	0.05	4.6	99.7	70.5
		0.05	0.05	0.1	4.6	96.9	53.5
		0.05	0.1	0.1	4.9	94.8	47.1
	25	0.01	0.01	0.05	4.6	94.6	50.9
		0.01	0.05	0.05	4.6	90.0	42.4
		0.05	0.05	0.05	4.9	89.5	42.0
		0.05	0.05	0.1	5.1	75.8	29.7
		0.05	0.1	0.1	4.7	71.6	26.7
15	100	0.01	0.01	0.05	4.7	100	99.9
		0.01	0.05	0.05	4.7	100	99.3
		0.05	0.05	0.05	4.8	100	99.1
		0.05	0.05	0.1	4.7	100	94.3
		0.05	0.1	0.1	4.6	100	90.9
	50	0.01	0.01	0.05	4.8	100	93.8
		0.01	0.05	0.05	4.7	100	87.6
		0.05	0.05	0.05	4.7	100	86.3
		0.05	0.05	0.1	4.7	99.8	69.8
		0.05	0.1	0.1	4.7	99.3	64.4
	25	0.01	0.01	0.05	4.7	99.3	68.0
		0.01	0.05	0.05	4.5	98.1	59.4
		0.05	0.05	0.05	4.5	97.7	56.9
		0.05	0.05	0.1	4.5	91.5	41.8
		0.05	0.1	0.1	4.5	87.5	37.7

本研究で誘導したWald-type検定統計量の限界として、欠測値を考慮した解析を行うことができないという問題がある。この問題については今後の更なる検討が必要である。また、提案する検定法は評価者のレベルが同程度であることを前提として誘導されていることから、評価者のレベルに差がある場合、特に二つの診断と評価者との間に質的交互作用が存在する場合には、提案する検定法では適切な評価ができないという問題がある。これについては、評価基準に対するトレーニングを全ての評価者に対して事前に実施し、評価レベルの統一化を図ることで解決できると考える。

参考文献

[1] Obuchowski NA. On the comparison of correlated proportions for clustered data. *Statistics in Medicine* 1998; **17**:1495–1507. DOI: 10.1002/(SICI)1097-0258(19980715)17:13<1495::AID-SIM863>3.0.CO;2-I

[2] Durkalski VL, Palesch YY, Lipsitz SR, Rust PF. Analysis of clustered matched-pair data for a non-inferiority study design. *Statistics in Medicine* 2003; **22**:279–290. DOI: 10.1002/sim.1385

[3] Nam J, Kwon D. Non-inferiority tests for clustered matched-pair data. *Statistics in Medicine* 2009; **28**:1668–1679. DOI: 10.1002/sim.3580

[4] Jin H, Lu Y. Comparison of correlated proportions based on paired binary data from clustered samples. *Journal of Statistical Planning and Inference* 2009; **139**:4206–4212. DOI: 10.1016/j.jspi.2009.06.005

[5] Lehr RG, Kashanian FK. Three persistent issues in analysis of clinical trials involving diagnostic contrast agents. *Drug Information Journal* 2009; **43**:525–532.

[6] Guidance for industry. Developing medical imaging drugs and biological products. Part 3: design, analysis, and interpretation of clinical studies, June 2004.
Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071604.pdf[16Apr2012].

[7] Obuchowski NA, Lieber ML. Statistics and methodology. *Skeletal Radiology* 2008; **37**:393–396. DOI: 10.1007/s00256-008-0448-1

[8] Appendix 1 to the guideline on clinical evaluation of diagnostic agents (CPMP/EWP/1119/98 REV. 1) on imaging agents (Doc. Ref. EMEA/CHMP/EWP/321180/2008), July 2009.
Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003581.pdf[16Apr2012].

[9] Saeki H, Tango T. Non-inferiority test and confidence interval for the difference in correlated proportions in diagnostic procedures based on multiple raters. *Statistics in Medicine* 2011; **30**:3313–3327. DOI: 10.1002/sim.4364