

두 진단검사의 비교에 대한 민감도와 특이도의 다변량 메타분석법

남선영^a, 송혜향^{1,a}

“가톨릭대학교 의생명과학교실 의학통계학과

요약

질병에 대한 새로운 진단검사 방법이 의학 연구자들에 의해 끊임없이 개발되고 있으며, 기존 진단검사 방법과 새로운 진단검사 방법을 비교하는 연구논문이 계속 출간되어 누적되고 있다. 메타분석법으로 다수 연구논문의 결과를 종합하여 정확성이 높은 진단검사에 대해 객관적인 결론을 내리게 된다. 이와같이 출간된 두 진단검사를 비교하는 각 연구논문은 각각 질병을 가진 개체와 질병을 가지지 않은 개체에 두 검사를 모두 실시하여 한 쌍의 민감도와 특이도를 구하여 비교한다. 이러한 연구논문의 결과를 종합하는 메타분석은 동일 개체에 실시한 두 검사로 인해 한 쌍의 민감도간의 연관성과 한 쌍의 특이도 간의 연관성을 고려한 메타분석법을 본 논문에서 제시한다. 논문예제 자료와 모의시험으로 메타분석 검정통계량의 효율성을 평가한다.

주요용어: 메타분석, 진단검사, 민감도와 특이도, 일반화 선형 혼합모형.

1. 서론

메타분석은 과거에 출간된 연구논문의 결과를 요약하고 재분석하는 통계적 방법이다. 본 논문에서는 진단검사를 비교하는 연구논문의 메타분석 방법론을 다루며 특히 상관성을 고려한 다변량 반응변수의 메타분석 방법론을 제시한다. 두 진단검사의 비교는 각각 질병군과 비질병군의 개체에게서 구해지는 민감도(Sensitivity)와 특이도(Specificity)에 근거하며, 또한 동일 대상자에 두 진단검사를 모두 실시하여 두 진단검사를 비교하는 일상적인 방법에서는 두 진단검사서 구해진 한 쌍의 특이도와 한 쌍의 민감도의 분석이 초점이다. 따라서 두 진단검사를 비교하는 연구논문의 메타분석에서는 다변량 변수의 연관성을 감안한 메타분석법의 제시가 중요하다.

질병군과 비질병군의 개체에게서 진단결과가 연속수치 또는 순위로 측정된 경우에 동일 경계값(threshold value)에 의해 민감도와 특이도가 결정되므로 동일 연구의 민감도와 특이도간에는 역상관관계가 관측되며 이를 무시한 메타분석법은 부적절하다고 한다 (Reitsma 등, 2005; Harbord 등, 2007). 이와 다르게 동일 대상자에게 두 진단검사를 모두 실시하여 구해진 한 쌍의 민감도 수치간에는 양의 상관성이 높을 것이 예상되고 또한 한 쌍의 특이도 수치간에도 높은 양의 상관성이 예상되며, 특히 이 두 상관성 수치가 서로 다른 경우를 임상에서 찾아 볼 수 있다. 따라서 이러한 다변량 변수간의 상관성을 감안하지 않은 메타분석은 부적절하며 이러한 이유로 다변량 메타분석법을 제안한다. 한편, 민감도와 특이도의 두 척도로부터 구한 ROC(Receiver Operating Characteristic) 곡선을 요약하고 병합하는 메타분석법은 민감도와 특이도 각각을 구분할 수 없는 방법이므로 본 논문에서는 고려하지 않는다 (Reitsma 등, 2005; Harbord 등, 2007).

본 논문에서는 관상동맥질환(coronary artery disease)의 진단에 사용되는 약물학적 압력 초음파 심장검진기(pharmacological stress echocardiography)에서 사용되는 두 압력측정법을 비교한 Picano 등

¹ 교신저자: (137-701) 서울 서초구 반포동 505 가톨릭대학교 대학원, 의생명과학교실 의학통계학과, 교수.
E-mail: hhsong@catholic.ac.kr

표 1: Dipyridamole-stress versus dobutamine-stress echocardiography for detection of coronary artery disease

| Authors | Year | Sensitivity | | Specificity | |
|-------------------------|------|--------------|-------------|--------------|------------|
| | | Dipyridamole | Dobutamine | Dipyridamole | Dobutamine |
| Martin <i>et al.</i> | 1992 | 14/25 (56)+ | 19/25 (76) | 5/9 (56) | 3/9 (33) |
| Salustri <i>et al.</i> | 1992 | 18/28 (64) | 16/28 (57) | 16/18 (89) | 14/18 (78) |
| Previtali <i>et al.</i> | 1993 | 34/57 (60) | 45/57 (79) | 22/23 (96) | 19/23 (83) |
| Beleslin <i>et al.</i> | 1994 | 88/119 (74) | 98/119 (82) | 16/17 (94) | 13/17 (76) |
| Gruber <i>et al.</i> | 1994 | 12/17 (70) | 12/17 (70) | 22/24 (92) | 22/24 (92) |
| Dagianti <i>et al.</i> | 1995 | 13/25 (52) | 18/25 (72) | 34/35 (97) | 34/35 (97) |
| Sochowski <i>et al.</i> | 1995 | 16/24 (67) | 17/24 (71) | 19/22 (86) | 18/22 (82) |
| Pingitore <i>et al.</i> | 1996 | 75/92 (82) | 77/92 (84) | 17/18 (94) | 15/18 (89) |
| San Roman <i>et al.</i> | 1996 | 49/63 (78) | 49/63 (78) | 38/39 (97) | 37/39 (95) |
| Minardi <i>et al.</i> | 1997 | 32/44 (73) | 33/44 (75) | 2/3 (67) | 2/3 (67) |
| Santoro <i>et al.</i> | 1998 | 18/33 (55) | 20/33 (60) | 26/27 (96) | 26/27 (96) |
| Battle <i>et al.</i> | 1998 | 34/41 (83) | 33/41 (80) | 15/15 (100) | 14/15 (93) |

+: Number of test positive/total number(%)

(2000)에 제시된 12편의 연구논문 자료에 다변량 메타분석법을 적용하여 결과를 요약하며, 단변량 메타분석법으로 요약한 메타분석 결과와 비교한다. 두 검사는 압력측정의 차이로써 첫 번째 검사는 Dipyridamole 압력측정법이며, 두 번째 검사는 Dobutamine 압력측정법이다. 표 1에 12편 연구논문의 각 검사결과인 민감도와 특이도가 제시되었다.

2. 상관성의 검토

표 1에 제시된 12편 논문의 동일 개체에 실시한 두 검사방법인 Dipyridamole(간략히 DI) 검사와 Dobutamine(간략히 DO) 검사의 민감도와 특이도로부터 계산한 공분산과 상관성을 살펴본다. 위의 행렬은 민감도와 특이도 수치의 공분산이고, 아래의 행렬은 로짓변환(logit transformation)된 민감도와 특이도 수치의 공분산이다. 아래 결과에서 A 는 민감도, B 는 특이도, 1은 DI 검사, 2는 DO 검사를 지칭한다.

$$\mathbf{cov} = \begin{pmatrix} \sigma_{A1}^2 & \sigma_{A1A2} & \sigma_{A1B1} & \sigma_{A1B2} \\ \sigma_{A2A1} & \sigma_{A2}^2 & \sigma_{A2B1} & \sigma_{A2B2} \\ \sigma_{B1A1} & \sigma_{B1A2} & \sigma_{B1}^2 & \sigma_{B1B2} \\ \sigma_{B2A1} & \sigma_{B2A2} & \sigma_{B2B1} & \sigma_{B2}^2 \end{pmatrix} = \begin{pmatrix} 0.011 & 0.005 & 0.003 & 0.003 \\ 0.005 & 0.007 & 0.0001 & -0.002 \\ 0.003 & 0.0001 & 0.018 & 0.022 \\ 0.003 & -0.002 & 0.022 & 0.032 \end{pmatrix},$$

$$\mathbf{logit(cov)} = \begin{pmatrix} 0.248 & 0.123 & 0.103 & 0.016 \\ 0.123 & 0.170 & 0.056 & -0.090 \\ 0.103 & 0.056 & 1.217 & 1.145 \\ 0.016 & -0.090 & 1.145 & 1.429 \end{pmatrix},$$

공분산 결과를 보면 DI 검사에서 민감도와 특이도간의 공분산은 0.003이고 DO 검사에서 민감도와 특이도간의 공분산은 -0.002이다. 그리고 민감도에서 두 검사간의 공분산은 0.005이고 특이도에서 두 검사간의 공분산은 0.022로 두 검사의 특이도간의 공분산이 매우 크게 나타났다. 분산을 살펴보면 각 검사에서 민감도의 분산 0.011과 0.007보다 각 검사에서 특이도의 분산 0.018과 0.032가 더욱 크게 나타났다. 로짓변환 후의 결과도 비슷한 경향이다.

다음으로 Pearson의 표본 상관계수를 살펴본다. 행렬의 대각선 위의 상관성 수치는 민감도와 특이

도 수치의 상관계수이며 대각선 아래의 수치는 로짓변환 된 민감도와 특이도 수치의 상관계수이다.

$$R = \begin{pmatrix} 1 & \rho_{A1A2} & \rho_{A1B1} & \rho_{A1B2} \\ & 1 & \rho_{A2B1} & \rho_{A2B2} \\ & & 1 & \rho_{B1B2} \\ & & & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0.549 & 0.206 & 0.186 \\ 0.596 & 1 & 0.006 & -0.128 \\ 0.187 & 0.122 & 1 & 0.916 \\ 0.028 & -0.182 & 0.868 & 1 \end{pmatrix},$$

이 상관계수의 유의성은 $t = r\sqrt{n-2}/\sqrt{1-r^2}$ 가 자유도 $n - 2$ 인 t 분포함을 이용하여 구하였다 (Snedecor와 Cochran, 1989, p.187). 여기서 n 은 연구논문 수이다.

$$P\text{-value} = \begin{pmatrix} \cdot & 0.06 & 0.52 & 0.56 \\ 0.04 & \cdot & 0.98 & 0.69 \\ 0.56 & 0.70 & \cdot & < 0.0001 \\ 0.93 & 0.57 & 0.0002 & \cdot \end{pmatrix},$$

로짓변환 전의 결과를 보면 DI 검사(검사 1)에서 민감도와 특이도의 상관계수(유의확률)이 0.206($P = 0.52$)이며 DO 검사(검사 2)에서 민감도와 특이도간의 상관계수는 $-0.128(P = 0.69)$ 로 모두 유의하지 않다. 그러나 DI 검사의 민감도와 DO 검사의 민감도간의 상관계수는 0.549($P = 0.06$)이며, DI 검사의 특이도와 DO 검사의 특이도간의 상관계수는 0.916($P < 0.0001$)으로 특이도간의 상관성은 민감도간의 상관성과 달리 매우 유의하다. 로짓변환 후의 민감도와 특이도의 결과를 살펴보면 로짓변환 전의 결과와 크게 다르지 않다.

종합하여 공분산과 표본상관계수의 결과로부터 민감도와 특이도간의 상관성은 매우 낮으며 그러나 한 쌍의 민감도간의 상관성 또는 한 쌍의 특이도간의 상관성은 매우 높다. 또한 한 쌍의 민감도간의 상관성과 한 쌍의 특이도간의 상관성의 정도가 다름을 볼 수 있으며 한 쌍의 특이도간의 상관성이 더욱 높게 나타났다.

한편, 두 민감도간과 두 특이도간의 양의 높은 상관성은 두 진단검사의 민감도의 비교와 또한 특이도의 비교에 대한 유의성 검정에 영향을 줄 수가 있으며, 논문 예제자료와 모의시험에서 이러한 영향을 알아보게 된다.

3. 메타분석 모형

여러 연구논문으로부터 동일 검사의 민감도와 특이도를 구했을 때 이변량 메타분석 모형으로 분석할 수도 있고, 또는 민감도와 특이도 각각을 단변량 모형으로 분석할 수도 있다. 한 단계 더 나아가 동일 개체에게 두 검사를 실시하여 두 검사결과로부터 한 쌍의 민감도와 한 쌍의 특이도를 구했을 때 다변량 모형으로 분석할 수도 있고 또는 각 검사를 독립되게 다루어 이변량 모형으로 분석할 수도 있다. 이와 같은 여러 분석모형을 설명하며, 랜덤효과 모형과 모수효과 모형을 간략히 설명한다. 한편 연구논문의 개체수가 충분히 크지 않거나, 민감도 또는 특이도가 0 또는 1에 가까운 값을 가지는 경우에는 로짓변환된 민감도와 로짓변환된 특이도의 정규분포 가정이 성립되지 않는다. 이러한 경우에는 비율에 근거한 랜덤효과 메타분석 모형을 이용할 수 있다 (Riley 등, 2007a; Cho와 Cole, 2006).

3.1. 이변량 랜덤효과 메타분석 모형

검사방법을 평가하려면 우선 기준 검사(reference test 또는 gold standard)에 의해 개체를 구분한다. 즉 기준 검사에서 검사결과가 양(+)으로 나타난 개체는 질병군으로, 음(-)으로 나타난 개체는 비질병군으로 분류된다. 이 기준 검사에 근거하여 다른 검사(index test)를 평가하며, i 번째 연구논문에서

표 2: 연구논문 한 편의 진단검사 결과

| Index test | Reference test | | |
|------------|----------------------|-----------------------------------------------------|-----------------------------------------------------|
| | Positive Negative | Positive | Negative |
| | | TP _i FN _i | FP _i TN _i |
| Total | | RP _i = TP _i + FN _i | RN _i = FP _i + TN _i |

총 개체수가 N_i 일 때, 참양성자수(true positive; TP), 위음성자수(false negative; FN), 위양성자수(false positive; FP)와 참음성자수(true negative; TN)로 판정된 개체수가 정해지면, 이로부터 특이도와 민감도가 구해진다. 즉, 민감도는 아래 표 2에서 양성자수 $RP_i = TP_i + FN_i$ 중에서 참양성자수 TP_i 의 비율이며, 특이도는 음성자수 $RN_i = FP_i + TN_i$ 중에서 참음성자수 TN_i 의 비율이다. 메타분석에서 총 연구논문수가 n 일 때, i 번째 ($i = 1, \dots, n$) 연구에서 관측된 민감도는 $P_{Ai} = TP_i/RP_i$ 이고, 관측된 특이도는 $P_{Bi} = TN_i/RN_i$ 로 정의된다.

이변량 랜덤효과 메타분석법(bivariate random-effects meta-analysis; BRMA)은 두 가지로 분류될 수 있으며, 각 검사방법에서 민감도와 특이도를 이변량으로 두어 분석하는 군간 이변량 랜덤효과 메타분석법(between-group BRMA; B-G BRMA)과 민감도와 특이도 각각에 대해 두 검사수치를 이변량으로 두어 분석하는 군내 이변량 랜덤효과 메타분석법(within-group BRMA; W-G BRMA)으로 구분된다. 군간 이변량 랜덤효과 메타분석(B-G BRMA)에서는 동일 개체에 적용한 두 검사의 한 쌍의 민감도는 서로 독립이고 또한 한 쌍의 특이도는 서로 독립이라는 가정에 근거하고, 군내 이변량 랜덤효과 메타분석(W-G BRMA)에서는 질병군과 비질병군이 서로 독립이라는 가정에 근거한다. 군간 이변량 모형(B-G BRMA)과 군내 이변량 모형(W-G BRMA)은 반응변수가 서로 다를 뿐 검정통계량의 형태가 동일하다.

이제 i 번째 연구의 두 검사방법 ($j = 1, 2$) 각각에서 로짓변환된 민감도와 로짓변환된 특이도를 각각 Y_{Aji} 와 Y_{Bji} 라 할 때 이에 근거한 군간 이변량 랜덤효과 메타분석(B-G BRMA) 모형은 다음과 같다.

$$\begin{aligned} \begin{pmatrix} Y_{Aji} \\ Y_{Bji} \end{pmatrix} &\sim N \left[\begin{pmatrix} \theta_{Aji} \\ \theta_{Bji} \end{pmatrix}, \delta_{ji} \right], \quad \delta_{ji} = \begin{pmatrix} s_{Aji}^2 & s_{Aji}s_{Bji}\rho_{wji} \\ s_{Aji}s_{Bji}\rho_{wji} & s_{Bji}^2 \end{pmatrix}, \\ \begin{pmatrix} \theta_{Aji} \\ \theta_{Bji} \end{pmatrix} &\sim N \left[\begin{pmatrix} \beta_{Aj} \\ \beta_{Bj} \end{pmatrix}, \Omega_j \right], \quad \Omega_j = \begin{pmatrix} \tau_{Aj}^2 & \tau_{Aj}\tau_{Bj}\rho_{vj} \\ \tau_{Aj}\tau_{Bj}\rho_{vj} & \tau_{Bj}^2 \end{pmatrix}, \end{aligned} \quad (3.1)$$

여기서 각 연구에서 θ_{Aji} 와 θ_{Bji} 는 Y_{Aji} 와 Y_{Bji} 의 참값(true value)이고, s_{Aji} 와 s_{Bji} 는 표준오차이며 ρ_{wji} 는 동일 연구에서의 로짓변환 민감도와 로짓변환 특이도 Y_{Aji} 와 Y_{Bji} 간의 연구내(within-study) 상관계수이다. 랜덤효과 모형에서는 각 연구에서의 검사대상자의 연령이나 질병의 중증도에 따른 차이 또는 다른 여러 가지의 요인에 의한 차이, 다시 말하면 서로 다른 연구논문들의 이질성을 분포로써 반영하여 검사방법을 평가한다. 즉, θ_{Aji} 와 θ_{Bji} 는 각각 총평균(pooled value) β_{Aj} 와 β_{Bj} 를 가지며, 표준오차가 τ_{Aj} 와 τ_{Bj} 이고, 연구간(between-study) 상관계수는 ρ_{vj} 인 정규분포에서 추출되었다고 가정한다. 일반적으로 총 평균인 β_{Aj} 와 β_{Bj} 의 추정값에 근거하여 메타분석에서 결론을 내리게 된다. 앞에서 설명한 랜덤효과 메타분석에서 θ_{Aji} 와 θ_{Bji} 의 분포를 고려하지 않고 상수로 취급하게 되면 모수효과(fixed effect) 모형이 된다.

군간 이변량 랜덤효과 메타분석(B-G BRMA) 모형은 연구내 상관계수 ρ_{wji} 와 연구간 상관계수 ρ_{vj} 를 감안한 모형이다. 군간 이변량 랜덤효과 메타분석(B-G BRMA) 모형에서 모수의 추정량과 이에 대한 분산-공분산 행렬은 다음과 같다. 두 검사방법 각각에서 민감도와 특이도를 비교하며, 검사 1과

검사 2는 독립이다. 즉, 아래 식 (3.2)에서 j 는 검사 1 또는 검사 2가 된다.

$$\hat{\beta}_j = \begin{pmatrix} \hat{\beta}_{Aj} \\ \hat{\beta}_{Bj} \end{pmatrix} = \left(\sum_{i=1}^n (\mathbf{\Omega}_j + \delta_{ji})^{-1} \right)^{-1} \left(\sum_{i=1}^n (\mathbf{\Omega}_j + \delta_{ji})^{-1} \mathbf{Y}_{ji} \right),$$

$$\text{cov}(\hat{\beta}_j) = \left(\sum_{i=1}^n (\mathbf{\Omega}_j + \delta_{ji})^{-1} \right)^{-1}, \quad (3.2)$$

여기서 $\mathbf{Y}_{ji} = (Y_{Aji}, Y_{Bji})'$ 이며, δ_{ji} 와 $\mathbf{\Omega}_j$ 는 식 (3.1)에 제시된 바와 같다. 또한 추정된 모수의 유의성 검정은 다음의 F 검정통계량을 이용한다. 여기서 n 은 연구논문수이다.

$$F = (\hat{\beta}_j)' \text{cov}(\hat{\beta}_j)^{-1} (\hat{\beta}_j) \sim F(1, n-2).$$

다음으로 군내 이변량 랜덤효과 메타분석(W-G BRMA) 모형은 i 번째 연구에서 민감도와 특이도($k = A, B$) 각각에서 두 검사에 대해 로짓변환된 값을 각각 Y_{k1i} 와 Y_{k2i} 라 할 때 다음과 같다.

$$\begin{pmatrix} Y_{k1i} \\ Y_{k2i} \end{pmatrix} \sim N \left[\begin{pmatrix} \theta_{k1i} \\ \theta_{k2i} \end{pmatrix}, \delta_{ki} \right], \quad \delta_{ki} = \begin{pmatrix} s_{k1i}^2 & s_{k1i}s_{k2i}\rho_{wki} \\ s_{k1i}s_{k2i}\rho_{wki} & s_{k2i}^2 \end{pmatrix},$$

$$\begin{pmatrix} \theta_{k1i} \\ \theta_{k2i} \end{pmatrix} \sim N \left[\begin{pmatrix} \beta_{k1} \\ \beta_{k2} \end{pmatrix}, \mathbf{\Omega}_k \right], \quad \mathbf{\Omega}_k = \begin{pmatrix} \tau_{k1}^2 & \tau_{k1}\tau_{k2}\rho_{vk} \\ \tau_{k1}\tau_{k2}\rho_{vk} & \tau_{k2}^2 \end{pmatrix}, \quad (3.3)$$

여기서 각 연구에서 θ_{k1i} 와 θ_{k2i} 는 Y_{k1i} 와 Y_{k2i} 의 참값(true value)이고, s_{k1i} 와 s_{k2i} 는 표준오차이며 ρ_{wki} 는 동일 연구에서의 두 검사에 대해 로짓변환된 값인 Y_{k1i} 와 Y_{k2i} 간의 연구내(within-study) 상관계수이다. 그리고 θ_{k1i} 와 θ_{k2i} 는 각각 총평균(pooled value) β_{k1} 와 β_{k2} 를 가지며, 표준오차가 τ_{k1} 와 τ_{k2} 이고, 연구 간(between-study) 상관계수는 ρ_{vk} 인 정규분포에서 추출되었다고 가정한다.

군내 이변량 랜덤효과 메타분석(W-G BRMA) 모형에서 모수의 추정량과 이에 대한 분산-공분산 행렬은 다음과 같다. 민감도와 특이도 각각에서 두 검사법을 비교하며, 민감도와 특이도는 독립이다. 즉, 아래 식 (3.4)에서 k 는 민감도 또는 특이도가 된다.

$$\hat{\beta}_k = \begin{pmatrix} \hat{\beta}_{k1} \\ \hat{\beta}_{k2} \end{pmatrix} = \left(\sum_{i=1}^n (\mathbf{\Omega}_k + \delta_{ki})^{-1} \right)^{-1} \left(\sum_{i=1}^n (\mathbf{\Omega}_k + \delta_{ki})^{-1} \mathbf{Y}_{ki} \right),$$

$$\text{cov}(\hat{\beta}_k) = \left(\sum_{i=1}^n (\mathbf{\Omega}_k + \delta_{ki})^{-1} \right)^{-1}, \quad (3.4)$$

여기서 $\mathbf{Y}_{ki} = (Y_{k1i}, Y_{k2i})'$ 이며, δ_{ki} 와 $\mathbf{\Omega}_k$ 는 식 (3.3)에 제시된 바와 같다. 또한 추정된 모수의 유의성 검정은 다음의 F 검정통계량을 이용한다. 여기서 n 은 연구논문수이다.

$$F = (\hat{\beta}_k)' \text{cov}(\hat{\beta}_k)^{-1} (\hat{\beta}_k) \sim F(1, n-2).$$

식 (3.1)과 (3.3)에서 s_{Aji} , s_{Bji} , s_{k1i} , s_{k2i} 와 ρ_{wji} , ρ_{wki} 는 알려져 있다고 가정하여 각 연구논문에서 이를 수집한다 (Van Hoawelingen 등, 2002). 특히 두 독립된 질병군과 비질병군으로부터 계산되는 민감도와 특이도의 경우에는 $\rho_{wji} = 0$ 으로 가정한다. 그러나 진단검사의 경우가 아닌 다수 반응변수에 경우에는 0으로 가정할 수 없다. 모수 $\beta_{Aj}\beta_{Bj}\tau_{Aj}\tau_{Bj}$ 의 추정은 제한된 일반화 최소자승법(Restricted Iterative Generalized Least Square; RIGLS)을 반복적으로 사용하여 추정하게 된다 (Riley 등, 2007b). 수렴된 추정결과의 확보에는 최소 4편 이상의 연구논문이 필요하다.

다변량 랜덤효과 메타분석(Univariate random-effect meta-analysis; URMA) 모형은 식 (3.3)에서 $\rho_{wki} = \rho_{vk} = 0$ 으로 둘 때이다. 그러므로 URMA 모형은 다음과 같다.

$$\begin{aligned} \begin{pmatrix} Y_{k1i} \\ Y_{k2i} \end{pmatrix} &\sim N \left[\begin{pmatrix} \theta_{k1i} \\ \theta_{k2i} \end{pmatrix}, \boldsymbol{\delta}_{ki} \right], \quad \boldsymbol{\delta}_{ki} = \begin{pmatrix} s_{k1i}^2 & 0 \\ 0 & s_{k2i}^2 \end{pmatrix}, \\ \begin{pmatrix} \theta_{k1i} \\ \theta_{k2i} \end{pmatrix} &\sim N \left[\begin{pmatrix} \beta_{k1} \\ \beta_{k2} \end{pmatrix}, \boldsymbol{\Omega}_k \right], \quad \boldsymbol{\Omega}_k = \begin{pmatrix} \tau_{k1}^2 & 0 \\ 0 & \tau_{k2}^2 \end{pmatrix}. \end{aligned} \quad (3.5)$$

3.2. 여러 검사 민감도와 특이도의 다변량 랜덤효과 메타분석 모형

본 논문의 예제 자료에서 고려하는 두 진단검사인 DI 검사와 DO 검사를 동일 개체에게 실시하여 각각 한 쌍의 민감도와 한 쌍의 특이도가 수집되는 4변량 자료의 다변량 랜덤효과 메타분석(multivariate random-effects meta-analysis; MRMA) 모형을 설명한다.

이제 i 번째($i = 1, \dots, n$) 연구에서 DI 검사(검사 1)와 DO 검사(검사 2)에서 각각의 민감도는 $P_{A1i} = TP_{1i}/RP_{1i}$ 와 $P_{A2i} = TP_{2i}/RP_{2i}$ 이고, 각각의 특이도는 $P_{B1i} = TN_{1i}/RN_{1i}$ 와 $P_{B2i} = TN_{2i}/RN_{2i}$ 이다. 또한 i 번째 연구에서 Y_{A1i} 와 Y_{A2i} 는 각각 로짓변환된 DI 검사와 DO 검사의 민감도이고, Y_{B1i} 와 Y_{B2i} 는 각각 로짓변환된 DI 검사와 DO 검사의 특이도이다. 각 연구에서 Y_{kji} ($k = A, B; j = 1, 2$)으로, Y_{kji} 와 θ_{kji} 의 MRMA 모형에서의 분포가정은 다음과 같다.

$$\begin{aligned} \begin{pmatrix} Y_{A1i} \\ Y_{A2i} \\ Y_{B1i} \\ Y_{B2i} \end{pmatrix} &\sim N \left[\begin{pmatrix} \theta_{A1i} \\ \theta_{A2i} \\ \theta_{B1i} \\ \theta_{B2i} \end{pmatrix}, \boldsymbol{\delta}_i \right], \quad \boldsymbol{\delta}_i = \begin{pmatrix} s_{A1i}^2 & s_{A1i}s_{A2i}\rho_{w_{A1A2}} & s_{A1i}s_{B1i}\rho_{w_{A1B1}} & s_{A1i}s_{B2i}\rho_{w_{A1B2}} \\ s_{A2i}s_{A1i}\rho_{w_{A2A1}} & s_{A2i}^2 & s_{A2i}s_{B1i}\rho_{w_{A2B1}} & s_{A2i}s_{B2i}\rho_{w_{A2B2}} \\ s_{B1i}s_{A1i}\rho_{w_{B1A1}} & s_{B1i}s_{A2i}\rho_{w_{B1A2}} & s_{B1i}^2 & s_{B1i}s_{B2i}\rho_{w_{B1B2}} \\ s_{B2i}s_{A1i}\rho_{w_{B2A1}} & s_{B2i}s_{A2i}\rho_{w_{B2A2}} & s_{B2i}s_{B1i}\rho_{w_{B2B1}} & s_{B2i}^2 \end{pmatrix}, \\ \begin{pmatrix} \theta_{A1i} \\ \theta_{A2i} \\ \theta_{B1i} \\ \theta_{B2i} \end{pmatrix} &\sim N \left[\begin{pmatrix} \beta_{A1} \\ \beta_{A2} \\ \beta_{B1} \\ \beta_{B2} \end{pmatrix}, \boldsymbol{\Omega} \right], \quad \boldsymbol{\Omega} = \begin{pmatrix} \tau_{A1}^2 & \tau_{A1}\tau_{A2}\rho_{v_{A1A2}} & \tau_{A1}\tau_{B1}\rho_{v_{A1B1}} & \tau_{A1}\tau_{B2}\rho_{v_{A1B2}} \\ \tau_{A2}\tau_{A1}\rho_{v_{A2A1}} & \tau_{A2}^2 & \tau_{A2}\tau_{B1}\rho_{v_{A2B1}} & \tau_{A2}\tau_{B2}\rho_{v_{A2B2}} \\ \tau_{B1}\tau_{A1}\rho_{v_{B1A1}} & \tau_{B1}\tau_{A2}\rho_{v_{B1A2}} & \tau_{B1}^2 & \tau_{B1}\tau_{B2}\rho_{v_{B1B2}} \\ \tau_{B2}\tau_{A1}\rho_{v_{B2A1}} & \tau_{B2}\tau_{A2}\rho_{v_{B2A2}} & \tau_{B2}\tau_{B1}\rho_{v_{B2B1}} & \tau_{B2}^2 \end{pmatrix}. \end{aligned} \quad (3.6)$$

다변량 랜덤효과 메타분석(MRMA) 모형에서 추정된 모수와 이에 대한 분산-공분산 행렬은 다음과 같다.

$$\begin{aligned} \hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_{A1} \\ \hat{\beta}_{A2} \\ \hat{\beta}_{B1} \\ \hat{\beta}_{B2} \end{pmatrix} &= \left(\sum_{i=1}^n (\boldsymbol{\Omega} + \boldsymbol{\delta}_i)^{-1} \right)^{-1} \left(\sum_{i=1}^n (\boldsymbol{\Omega} + \boldsymbol{\delta}_i)^{-1} \mathbf{Y}_i \right), \\ \text{cov}(\hat{\boldsymbol{\beta}}) &= \left(\sum_{i=1}^n (\boldsymbol{\Omega} + \boldsymbol{\delta}_i)^{-1} \right)^{-1}, \end{aligned} \quad (3.7)$$

여기서 $\mathbf{Y}_{ji} = (Y_{A1i}, Y_{A2i}, Y_{B1i}, Y_{B2i})'$ 이며, $\boldsymbol{\delta}_i$ 와 $\boldsymbol{\Omega}$ 는 식 (3.6)에 제시된 바와 같다.

추정된 모수의 유의성 검정은 다음과 같은 F 검정통계량을 이용한다. 여기서 n 은 연구논문수이며, l 은 검사의 가지수이다.

$$F = (\hat{\boldsymbol{\beta}})' \text{cov}(\hat{\boldsymbol{\beta}})^{-1} (\hat{\boldsymbol{\beta}}) \sim F(1, (n-2)(l-1)),$$

식 (3.6)의 모수 β_{kj} 와 τ_{kj}^2 ($j = 1, 2; k = A, B$)의 추정엔 제한된 일반화 최소자승법(Restricted Iterative Generalized Least Square; RIGLS)을 반복적으로 사용하여 추정하게 된다 (Riley 등, 2007b). 수렴된 추정결과와 확보에는 최소 8편 이상의 연구논문이 필요하다. 식 (3.6)에서 $\rho_{wi_{kj}} = \rho_{v_{kj}} = 0$ ($k = A, B; j = 1, 2$)으로 둘 때 네 개의 독립인 URMA 모형이 된다.

$$\begin{pmatrix} Y_{A1i} \\ Y_{A2i} \\ Y_{B1i} \\ Y_{B2i} \end{pmatrix} \sim N \left[\begin{pmatrix} \theta_{A1i} \\ \theta_{A2i} \\ \theta_{B1i} \\ \theta_{B2i} \end{pmatrix}, \delta_i \right], \quad \delta_i = \begin{pmatrix} s_{A1i}^2 & 0 & 0 & 0 \\ 0 & s_{A2i}^2 & 0 & 0 \\ 0 & 0 & s_{B1i}^2 & 0 \\ 0 & 0 & 0 & s_{B2i}^2 \end{pmatrix},$$

$$\begin{pmatrix} \theta_{A1i} \\ \theta_{A2i} \\ \theta_{B1i} \\ \theta_{B2i} \end{pmatrix} \sim N \left[\begin{pmatrix} \beta_{A1} \\ \beta_{A2} \\ \beta_{B1} \\ \beta_{B2} \end{pmatrix}, \Omega \right], \quad \Omega = \begin{pmatrix} \tau_{A1}^2 & 0 & 0 & 0 \\ 0 & \tau_{A2}^2 & 0 & 0 \\ 0 & 0 & \tau_{B1}^2 & 0 \\ 0 & 0 & 0 & \tau_{B2}^2 \end{pmatrix}, \quad (3.8)$$

단변량 랜덤효과 메타분석(URMA) 모형에서 j 번째 검사의 민감도($k = A$)와 특이도($k = B$)의 추정량과 분산은 구체적으로 다음과 같다. 여기서 각 검사에 대한 민감도와 특이도는 모두 독립이다.

$$\hat{\beta}_{kj} = \frac{\sum_{i=1}^n Y_{kji} / (s_{kji}^2 + \hat{\tau}_{kj}^2)}{\sum_{i=1}^n 1 / (s_{kji}^2 + \hat{\tau}_{kj}^2)} = \frac{\sum_{i=1}^n w_{kji} Y_{kji}}{\sum_{i=1}^n w_{kji}},$$

$$\text{var}(\hat{\beta}_{kj}) = \frac{1}{\sum_{i=1}^n 1 / (s_{kji}^2 + \hat{\tau}_{kj}^2)} = \frac{1}{\sum_{i=1}^n w_{kji}}, \quad (3.9)$$

여기서

$$w_{kji} = \frac{1}{s_{kji}^2 + \hat{\tau}_{kj}^2}$$

이다. 추정된 모수의 유의성 검정은 다음의 t 검정통계량을 이용한다. 여기서 n 은 연구논문수이다.

$$t = \frac{\hat{\beta}_{kj}}{\text{SE}(\hat{\beta}_{kj})} \sim t(n-1)$$

그리고 식 (3.6)에서 $\rho_{wi_{A1B1}} = \rho_{wi_{A1B2}} = \rho_{wi_{A2B1}} = \rho_{wi_{A2B2}} = 0$ 으로 둘 때 W-G BRMA 모형이 되며, $\rho_{wi_{A1A2}} = \rho_{wi_{A2A1}} = \rho_{wi_{B1B2}} = \rho_{wi_{B2B1}} = 0$ 으로 둘 때 B-G BRMA 모형이 된다.

4. 예제자료 분석

1장에 제시된 관상동맥질환의 두 가지 진단검사에 대해 12편 연구논문 자료를 Menke (2010)에 제시되어 있는 프로그램을 참고하여 분석한다. Menke (2010)는 한 가지 진단방법에 대한 민감도와 특이도의 이변량 메타분석에 대하여 프로그램을 제시하였으며, 이에 바탕하여 다변량 메타분석으로 확장하였다.

4.1. 단변량 랜덤효과 메타분석(URMA)

단변량 랜덤효과 메타분석에 의해 추정된 모수 $\hat{\beta}_{A1}$, $\hat{\beta}_{A2}$, $\hat{\beta}_{B1}$, $\hat{\beta}_{B2}$ 와 모수추정량의 표준오차가 표 3에 제시되었다. 그리고 이 추정된 모수로부터 민감도와 특이도의 추정량 \hat{P} 와 \hat{P} 의 표준오차가 제시되

표 3: URMA 모형에 의한 모수추정량과 분산

| | 검사 | $\hat{\beta}$ | $\hat{\beta}$ 의 표준오차(SE) | $\hat{P} = (e^{\hat{\beta}})/(1 + e^{\hat{\beta}})$ | \hat{P} 의 표준오차(SE) |
|-----|----|---------------|--------------------------|-----------------------------------------------------|----------------------|
| 민감도 | DI | 0.815 | 0.149 | 0.693 | 0.155 |
| | DO | 1.140 | 0.129 | 0.758 | 0.154 |
| 특이도 | DI | 2.484 | 0.329 | 0.923 | 0.168 |
| | DO | 1.787 | 0.324 | 0.857 | 0.168 |

표 4: W-G BRMA 모형에 의한 모수추정량과 분산

| | 검사 | $\hat{\beta}$ | $\hat{\beta}$ 의 표준오차(SE) | $\hat{P} = (e^{\hat{\beta}})/(1 + e^{\hat{\beta}})$ | \hat{P} 의 표준오차(SE) |
|-----|----|---------------|--------------------------|-----------------------------------------------------|----------------------|
| 민감도 | DI | 0.785 | 0.150 | 0.687 | 0.155 |
| | DO | 1.121 | 0.135 | 0.754 | 0.154 |
| 특이도 | DI | 2.490 | 0.371 | 0.923 | 0.171 |
| | DO | 1.773 | 0.346 | 0.855 | 0.169 |

었다. DI와 DO 검사 모두에서 URMA 모형에 의해 추정된 민감도보다 특이도가 더욱 크며, 표준오차도 더욱 크다. URMA 모형에 의해 추정된 연구간 분산-공분산 행렬은 다음과 같다.

$$\Omega = \begin{pmatrix} \tau_{A1}^2 & 0 & 0 & 0 \\ 0 & \tau_{A2}^2 & 0 & 0 \\ 0 & 0 & \tau_{B1}^2 & 0 \\ 0 & 0 & 0 & \tau_{B2}^2 \end{pmatrix} = \begin{pmatrix} 0.144 & 0 & 0 & 0 \\ 0 & 0.069 & 0 & 0 \\ 0 & 0 & 0.522 & 0 \\ 0 & 0 & 0 & 0.767 \end{pmatrix}$$

DI 검사의 특이도의 분산은 0.522이고 DO 검사의 특이도의 분산은 0.767로 민감도의 경우보다 매우 큰 것으로 나타났다.

4.2. 이변량 랜덤효과 메타분석(BRMA)

앞의 2장에서 살펴본 상관성은 군내 이변량 모형(W-G BRMA)을 뒷받침하고 있고 또한 질병군과 비질병군이 서로 독립된 개체이므로 본 논문에서는 군내 이변량 모형(W-G BRMA)을 중점으로 다루며, 다변량 랜덤효과 메타분석(MRMA)에서는 군내 이변량 모형(W-G BRMA)과 군간 이변량 모형(B-G BRMA)을 모두 감안하여 분석한다. 이제 두 검사의 한 쌍의 민감도간에 연관성을 고려하고 마찬가지로 두 검사의 한 쌍의 특이도 간의 연관성을 고려한 군내 이변량 랜덤효과 모형(W-G BRMA)의 분석 결과는 다음과 같다.

W-G BRMA 모형에 의해 추정된 값이 URMA 모형의 경우에 비해 약간 작은 값을 가지나, 추정량의 분산은 약간 커졌다. W-G BRMA 모형에 의해 추정된 연구간 분산-공분산 행렬은 다음과 같다.

$$\Omega = \begin{pmatrix} \tau_{A1}^2 & \tau_{A1}\tau_{A2}\rho_{v_{A1A2}} & 0 & 0 \\ \tau_{A1}\tau_{A2}\rho_{v_{A1A2}} & \tau_{A2}^2 & 0 & 0 \\ 0 & 0 & \tau_{B1}^2 & \tau_{B1}\tau_{B2}\rho_{v_{B1B2}} \\ 0 & 0 & \tau_{B1}\tau_{B2}\rho_{v_{B1B2}} & \tau_{B2}^2 \end{pmatrix} = \begin{pmatrix} 0.152 & 0.117 & 0 & 0 \\ 0.117 & 0.091 & 0 & 0 \\ 0 & 0 & 0.854 & 0.897 \\ 0 & 0 & 0.897 & 0.943 \end{pmatrix}$$

한 쌍의 특이도간의 공분산이 0.897로 큰 값을 가지며 이러한 결과는 모든 특이도와 민감도를 독립되게 다룬 4.1절의 URMA 모형의 분산-공분산 행렬이 적절하지 않고, 따라서 URMA 모형이 옳지 않음을 시사한다.

4.3. 다변량 랜덤효과 메타분석(MRMA)

다변량 랜덤효과 메타분석에 의해 추정된 모수 $\hat{\beta}_{A1}$, $\hat{\beta}_{A2}$, $\hat{\beta}_{B1}$, $\hat{\beta}_{B2}$ 와 추정된 모수의 표준오차가 표

표 5: MRMA 모형에 의한 모수추정량과 분산

| | 검사 | $\hat{\beta}$ | $\hat{\beta}$ 의 표준오차(SE) | $\hat{P} = (e^{\hat{\beta}})/(1 + e^{\hat{\beta}})$ | \hat{P} 의 표준오차(SE) |
|-----|----|---------------|--------------------------|-----------------------------------------------------|----------------------|
| 민감도 | DI | 0.778 | 0.149 | 0.685 | 0.155 |
| | DO | 1.115 | 0.136 | 0.753 | 0.154 |
| 특이도 | DI | 2.466 | 0.358 | 0.922 | 0.170 |
| | DO | 1.752 | 0.340 | 0.852 | 0.169 |

표 6: 두 검사간의 차이에 대한 검정결과인 P-value

| | 민감도에서 두 검사의 차이 | 특이도에서 두 검사의 차이 |
|----------|----------------|----------------|
| URMA | 0.129 | 0.162 |
| W-G BRMA | 0.037 | 0.049 |
| MRMA | 0.018 | 0.028 |

5에 제시되어 있다. 그리고 이 추정된 모수로부터 민감도와 특이도의 추정량 \hat{P} 와 \hat{P} 의 표준오차가 제시되어 있다.

MRMA 모형에 의해 추정된 값이 URMA 모형과 BRMA 모형의 경우에 비해 더욱 작은 값을 가지며 그러나 분산은 URMA 모형의 경우보다 약간 크게 나타났다. MRMA 모형에 의해 추정된 연구간 분산-공분산 행렬과 상관계수 행렬은 다음과 같다.

$$\begin{aligned}
 \mathbf{\Omega} &= \begin{pmatrix} \tau_{A1}^2 & \tau_{A1}\tau_{A2}\rho_{VA1A2} & \tau_{A1}\tau_{B1}\rho_{VA1B1} & \tau_{A1}\tau_{B2}\rho_{VA1B2} \\ \tau_{A2}\tau_{A1}\rho_{VA2A1} & \tau_{A2}^2 & \tau_{A2}\tau_{B1}\rho_{VA2B1} & \tau_{A2}\tau_{B2}\rho_{VA2B2} \\ \tau_{B1}\tau_{A1}\rho_{VB1A1} & \tau_{B1}\tau_{A2}\rho_{VB1A2} & \tau_{B1}^2 & \tau_{B1}\tau_{B2}\rho_{VB1B2} \\ \tau_{B2}\tau_{A1}\rho_{VB2A1} & \tau_{B2}\tau_{A2}\rho_{VB2A2} & \tau_{B2}\tau_{B1}\rho_{VB2B1} & \tau_{B2}^2 \end{pmatrix} \\
 &= \begin{pmatrix} 0.150 & 0.113 & 0.107 & 0.025 \\ 0.113 & 0.093 & 0.008 & -0.064 \\ 0.107 & 0.008 & 0.763 & 0.805 \\ 0.025 & -0.064 & 0.805 & 0.906 \end{pmatrix} \\
 \mathbf{R} &= \begin{pmatrix} 1 & 0.957 & 0.316 & 0.068 \\ 0.957 & 1 & 0.030 & -0.220 \\ 0.316 & 0.030 & 1 & 0.968 \\ 0.068 & -0.220 & 0.968 & 1 \end{pmatrix}
 \end{aligned}$$

다변량 랜덤효과 모형으로 추정된 상관계수를 살펴보면 한 쌍의 민감도간의 상관계수와 한 쌍의 특이도간의 상관계수가 매우 높으며 또한 민감도와 특이도간에도 결코 0이 아닌 상관계수 결과를 볼 수 있다. 즉, 다변량 랜덤효과 메타분석에 의해 분석되어야 함을 뜻한다. 2절에서 계산한 표본상관계수와 비교하였을 때 로짓변환 후 DO 검사의 민감도와 DI 검사의 특이도의 상관계수를 제외하고 앞의 2절보다 상관계수가 크게 추정되었음을 알 수 있다.

두 검사간의 차이가 있는지에 대한 검정, 즉 두 검사의 민감도가 서로 다르고 또한 두 검사의 특이도가 다른가에 대한 검정결과가 표 6에 제시되었다. 두 검사간의 차이에 대하여 우도비 검정을 사용하였다. 네 변수를 모두 독립으로 보는 경우 URMA 모형에서는 민감도에서 두 검사와 특이도에서 두 검사가 차이가 없는 것으로 나타났다. 그러나 두 검사의 민감도간의 연관성과 두 검사의 특이도간의 연관성을 감안한 W-G BRMA 모형과 모든 수치간의 연관성을 감안한 MRMA 모형에서는 두 검사의 민감도와 특이도가 각각 모두 유의한 것으로 나타났다. 특히 MRMA 모형에서 더욱 유의하다.

5. 모의시험

5.1. 계획

모의시험의 목적은 URMA, W-G BRMA, MRMA 세 가지 분석법을 비교하는 것이다. 모의시험의 가설은 다음과 같다.

H_0 : 민감도(특이도)에서 두 검사의 차이가 없다.

H_1 : 민감도(특이도)에서 두 검사의 차이가 있다.

우선 먼저 각 검사에 대한 민감도와 특이도의 4변량 자료를 다음과 같이 두 가지 모형에 따라 생성한다. 자료는 프로그램 R로 생성하였다.

모형 (1)

$$\begin{pmatrix} \theta_{A1i} \\ \theta_{A2i} \\ \theta_{B1i} \\ \theta_{B2i} \end{pmatrix} \sim N \left[\begin{pmatrix} \beta_{A1} \\ \beta_{A2} \\ \beta_{B1} \\ \beta_{B2} \end{pmatrix}, \mathbf{\Omega} \right], \quad \mathbf{\Omega} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

모형 (2)

$$\begin{pmatrix} \theta_{A1i} \\ \theta_{A2i} \\ \theta_{B1i} \\ \theta_{B2i} \end{pmatrix} \sim N \left[\begin{pmatrix} \beta_{A1} \\ \beta_{A2} \\ \beta_{B1} \\ \beta_{B2} \end{pmatrix}, \mathbf{\Omega} \right], \quad \mathbf{\Omega} = \begin{pmatrix} 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 \end{pmatrix},$$

여기서 모형 (1)은 각 검사에서 민감도와 특이도가 독립이라는 가정을 나타내고, 모형 (2)는 각 검사에서 민감도와 특이도의 공분산이 모두 동일한 상수값을 가짐을 나타낸다. 귀무가설하에서 β_{A1} , β_{A2} , β_{B1} , β_{B2} 은(logit(0.7), logit(0.7), logit(0.9), logit(0.9))으로 선택하고, 대립가설하에서 β_{A1} , β_{A2} , β_{B1} , β_{B2} 은(logit(0.7), logit(0.8), logit(0.8), logit(0.9))로 선택하였다.

한 연구에서 총 피험자의 수를 50명으로 고정하여 위에서 생성된 각 검사에 대한 민감도와 특이도를 이용하여 참양성자수(TP), 참음성자수(TN), 위양성자수(FP), 위음성자수(FN)을 계산한다. 연구의 수는 총 30편으로 하고 이 30편의 연구에 대하여 세 가지 메타분석을 실시한다. 이 모의시험을 SAS를 이용하여 총 1000번 시행하였다.

5.2. 결과

세 가지 분석에 대하여 추정된 모수와 모수에 대한 표준오차의 평균값을 비교한 표이다. 네 개의 변수가 모두 독립이라는 가정 하에서 세 가지 메타분석을 시행한 결과가 표 7에 나타나있다. 세 가지 방법에서 추정된 모수가 크게 차이가 없는 것으로 보인다.

네 개의 변수에 대한 공분산이 모두 동일하다는 가정 하에서 실시한 메타분석의 결과를 표 8에 나타내었다. 세 가지 방법에서 추정된 모수가 크게 차이가 없는 것으로 보인다. W-G BRMA와 MRMA 방법에서 추정된 값이 모수에 가장 가까운 값을 나타내었다.

표 9에는 세 가지 분석법에서 각 모형에 대하여 추정된 유의수준이 제시되었다. 우선 민감도에서 두 검사의 차이에 대한 유의수준을 비교해 보면 모든 변수가 독립이라는 가정 하에서 URMA 방법이 0.05에 가장 가까운 값을 나타냈다. 모든 변수의 공분산이 같다는 가정 하에서는 MRMA 방법에서

표 7: 모형 (1)에 의해 추정된 모수와 추정된 모수의 표준오차

| | 귀 무 가 설 | | | |
|----------|---------------|---------------|---------------|---------------|
| | 검사1 민감도 = 0.7 | 검사2 민감도 = 0.7 | 검사1 특이도 = 0.9 | 검사2 특이도 = 0.9 |
| URMA | 0.682(0.0541) | 0.690(0.0543) | 0.885(0.0540) | 0.870(0.0541) |
| W-G BRMA | 0.683(0.0542) | 0.690(0.0543) | 0.885(0.0539) | 0.870(0.0541) |
| MRMA | 0.683(0.0546) | 0.690(0.0546) | 0.886(0.0546) | 0.871(0.0547) |
| | 대 립 가 설 | | | |
| | 검사1 민감도 = 0.7 | 검사2 민감도 = 0.8 | 검사1 특이도 = 0.8 | 검사2 특이도 = 0.9 |
| URMA | 0.687(0.0543) | 0.778(0.0542) | 0.799(0.0541) | 0.869(0.0539) |
| W-G BRMA | 0.692(0.0543) | 0.779(0.0543) | 0.803(0.0541) | 0.869(0.0540) |
| MRMA | 0.688(0.0544) | 0.779(0.0545) | 0.803(0.0543) | 0.870(0.0543) |

표 8: 모형 (2)에 의해 추정된 모수와 추정된 모수의 표준오차

| | 귀 무 가 설 | | | |
|----------|---------------|---------------|---------------|---------------|
| | 검사1 민감도 = 0.7 | 검사2 민감도 = 0.7 | 검사1 특이도 = 0.9 | 검사2 특이도 = 0.9 |
| URMA | 0.694(0.0540) | 0.690(0.0541) | 0.882(0.0539) | 0.874(0.0541) |
| W-G BRMA | 0.694(0.0541) | 0.691(0.0542) | 0.883(0.0539) | 0.875(0.0541) |
| MRMA | 0.695(0.0554) | 0.691(0.0554) | 0.884(0.0567) | 0.876(0.0568) |
| | 대 립 가 설 | | | |
| | 검사1 민감도 = 0.7 | 검사2 민감도 = 0.8 | 검사1 특이도 = 0.8 | 검사2 특이도 = 0.9 |
| URMA | 0.705(0.0542) | 0.797(0.0543) | 0.801(0.0538) | 0.879(0.0538) |
| W-G BRMA | 0.706(0.0543) | 0.798(0.0543) | 0.802(0.0539) | 0.882(0.0542) |
| MRMA | 0.706(0.0559) | 0.798(0.0566) | 0.802(0.0556) | 0.882(0.0573) |

표 9: 세 가지 메타분석법의 제 1종 오류

| | | 민감도에서 두 검사의 차이 | | 특이도에서 두 검사의 차이 | |
|--------|----------|----------------|-------|----------------|-------|
| | | | | | |
| 모형 (1) | URMA | | 0.060 | | 0.068 |
| | W-G BRMA | | 0.061 | | 0.078 |
| | MRMA | | 0.061 | | 0.078 |
| 모형 (2) | URMA | | 0.031 | | 0.034 |
| | W-G BRMA | | 0.048 | | 0.063 |
| | MRMA | | 0.049 | | 0.058 |

표 10: 세 가지 메타분석법의 검정력

| | | 민감도에서 두 검사의 차이 | | 특이도에서 두 검사의 차이 | |
|--------|----------|----------------|-------|----------------|-------|
| | | | | | |
| 모형 (1) | URMA | | 0.499 | | 0.523 |
| | W-G BRMA | | 0.498 | | 0.524 |
| | MRMA | | 0.500 | | 0.533 |
| 모형 (2) | URMA | | 0.539 | | 0.658 |
| | W-G BRMA | | 0.776 | | 0.815 |
| | MRMA | | 0.781 | | 0.832 |

0.05에 가장 가까운 값을 나타냈다. 특이도에서 두 검사의 차이에 대한 유의수준을 비교해 보면 모든 변수가 독립이라는 가정 하에서는 URMA 방법이 0.05에 가장 가까운 값을 나타냈다. 모든 변수의 공분산이 같다는 가정한 모형에서는 MRMA 방법이 조금 더 0.05에 가까운 값을 나타냈다.

표 10에는 세 가지 분석법에서 각 모형에 대하여 검정력이 제시되었다. 우선 민감도에서 두 검사의 차이에 대한 검정력을 비교해 보면 모든 변수가 독립이라는 가정 하에서 MRMA 방법이 조금 더 높

은 검정력을 나타냈다. 모든 변수의 공분산이 같다는 가정 하에서는 URMA 방법에 비해서 나머지 두 방법이 훨씬 높은 검정력을 나타내고, MRMA 방법에서 가장 높은 검정력을 가진다. 특이도에서 두 검사의 차이에 대한 검정력을 비교해 보면 모든 변수가 독립이라는 가정 하에서는 세 방법이 비슷하지만, MRMA 방법이 조금 더 높은 검정력을 가진다. 모든 변수의 공분산이 같다는 가정된 모형에서는 URMA 방법보다 나머지 두 방법이 더 높은 검정력을 가지며, MRMA 방법이 가장 높은 검정력을 가진다.

6. 결론

두 진단검사의 평가에서는 민감도와 특이도를 동시에 분석해야 할 뿐만 아니라 동일 개체에게서 한 쌍의 민감도와 한 쌍의 특이도가 측정되는 반복측정 자료의 경우이다. 이러한 다변량 자료는 여러 횃수의 일변량 검정법을 적용하여서도 분석이 가능하다. 그러나 이와같은 분석법은 여러 횃수 검정의 적용으로 다중검정(multiplicity)의 문제점을 가지게 된다. 다시 말하면 제 1종의 오류가 매우 크게 되며 따라서 결과의 해석에 많은 주의가 요구된다. 다변량 분석법은 두 진단검사의 민감도와 특이도간의 연관된 자료의 특성을 분석과정에서 충실하게 감안하여 분석하게 되므로 결론에 대해 신뢰할 수 있다 하겠다. 다변량 랜덤효과 모형으로 분석되어야 하는 연구논문 자료를 단변량 랜덤효과 모형으로 분석하거나 또는 이변량 랜덤효과 모형으로 분석하게 되면 추정된 모수값이 과추정 될 수 있음을 볼 수 있으며, 다변량 랜덤효과 모형의 더욱 적절함은 추정된 공분산 행렬과 상관계수 행렬로써 확인할 수 있으며 어떠한 수치도 독립되게 다루어 질 수 없음을 알 수 있다.

참고 문헌

- Cho, 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**, 1331–1333.
- Harbord, R. M., Deeks, J. J., Egger, M., Whiting, P. and Sterne, J. A. C. (2007). A unification of models for meta-analysis of diagnostic accuracy studies, *Biostatistics*, **8**, 239–251.
- Menke, J. (2010). Bivariate random-effects meta-analysis of sensitivity and specificity with SAS PROC GLIMMIX, *Methods Inf Med*, **49**, 54–64.
- Picano, E., Bedetti, G., Varga, A. and Cseh, E. (2000). The comparable diagnostic accuracies of dobutamine-stress and dipyridamole-stress echocardiographies: A meta-analysis, *Coronary Artery Disease*, **11**, 151–159.
- Reitsma, J. B., Glas, A. S., Rutjes, A. W. S., Rob, J. S., 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**, 982–990.
- Riley, R. D., Abrams, K. R., Sutton, A. J., Lambert, P. C. and Thompson, J. R. (2007a). Bivariate random-effects meta-analysis and the estimation of between-study correlation, *BMC Medical Research Methodology*, **7**, 3.
- Riley, R. D., Abrams, K. R., Lambert, P. C., Sutton, A. J. and Thompson, J. R. (2007b). An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes, *Statistics in Medicine*, **26**, 78–97.
- Snedecor, G. W. and Cochran, W. G. (1989). *Statistical Methods, Eighth Edition*, Iowa State University Press.
- van Houwelingen, H. C., Arends, L. R. and Theo, S. (2002). Advanced methods in meta-analysis: multivariate approach and meta-regression, *Statistics in Medicine*, **21**, 589–624.

Multivariate Meta-Analysis Methods of Comparing the Sensitivity and Specificity of Two Diagnostic Tests

Seon-Young Nam^a, Hae-Hiang Song^{1,a}

^aBiostatistics Division, Department of Medical Lifescience, The Catholic University of Korea

Abstract

Researchers are continuously trying to find innovative diagnostic tests and published articles are accumulating at an enormous rate in many medical fields. Meta-analysis enables previously published study results to be reviewed and summarized; therefore, an objective assessment of diagnostic tests can be done with a meta-analysis of sensitivities and specificities. Data obtained by applying two diagnostic tests to a well-defined group of diseased patients produce a pair of sensitivity and by applying the same medical tests to a group of non-diseased subjects produce a pair of specificity. The statistical tests in the meta-analysis need to consider the correlatedness of the results from two diagnostic tests applied to the same diseased and non-diseased subjects. The associations between two diagnostic test results are often found to be unequal for the diseased and non-diseased subjects. In this paper, multivariate meta-analytic methods are studied by taking into account the different associations between correlated variables. On the basis of Monte Carlo simulations, we evaluate the performance of the multivariate meta-analysis methods proposed in this paper.

Keywords: Meta analysis, diagnostic test, sensitivity and specificity, generalized linear mixed model.

¹ Corresponding author: Professor, Biostatistics Division, Department of Medical Lifescience, Graduate School, The Catholic University of Korea, 505 Banpo-Dong, Seocho-Gu, Seoul 137-701, Korea. E-mail: hhsong@catholic.ac.kr