



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

여러 평가자들의 진단자료에서 질병 예측의
최적 기준점 선정에 대한 비교 연구

연세대학교 대학원

의학전산통계학협동과정

의학통계학전공

김 경 민

여러 평가자들의 진단자료에서 질병 예측의 최적 기준점 선정에 대한 비교 연구

지도 정 인 경 교수

이 논문을 석사 학위논문으로 제출함

2016년 12월 일

연세대학교 대학원

의학전산통계학협동과정

의학통계학전공

김 경 민

김경민의 석사 학위논문을 인준함

심사위원 정 인 경 인

심사위원 남 정 모 인

심사위원 한 경 화 인

연세대학교 대학원

2016년 12월 일

차 례

표 차례	iv
그림 차례	iv
국문 요약	v
제 1 장 서론	1
1.1 연구 배경 및 목적	1
1.2 연구 내용 및 방법	2
1.3 논문의 구성	3
제 2 장 이론적 배경	4
1. Pooled ROC Curve	4
2. 최적 기준점 선정 방법	5
2.1. Efficiency (EF)	5
2.2. W.J.Youden 의 J 통계량(Youden' s index)	6
2.3. 유클리디안(Euclidean)	7
2.4. 우도비(Likelihood Ratio)	9
2.5. 상호 정보(Mutual Information)	11
제 3 장 모의실험	16
3.1 모의실험 설계	17
3.2 모의실험 결과	24
3.2.1 평가자를 고정효과로 가정한 모형	26
3.2.2 평가자를 임의효과로 가정한 모형	27

제 4 장 실제 자료 적용	31
제 5 장 결론 및 고찰	33
참고문헌	36
부록	40
그림 A_1. Random effect(# of 3), Equal variance, Pattern=HL일 때, 최적 기준점(AUC=0.702, 0.855, 0.961)	40
그림 A_2. Random effect(# of 3), Equal variance, Pattern=LL일 때, 최적 기준점(AUC=0.702, 0.855, 0.961)	40
그림 A_3. Random effect(# of 3), Equal variance, Pattern=HH일 때, 최적 기준점(AUC=0.702, 0.855, 0.961)	41
그림 A_4. Random effect(# of 3), Equal variance, Pattern=LH일 때, 최적 기준점(AUC=0.702, 0.855, 0.961)	41
그림 B. Random effect(# of 3), Equal variance, Pattern=HH일 때, 두 방법(Youden index, Euclidean)의 최적 기준점과 실제 기준점 최적 기준점(AUC=0.702, 0.855, 0.961)	42
그림 C_1. Random effect(# of 3), Equal variance, Pattern=HH일 때, 최적 기준점(AUC=0.702, 0.855, 0.961)	43
그림 C_2. Random effect(# of 3), Unequal variance, Pattern=HH일 때, 최적 기준점(AUC=0.702, 0.855, 0.961)	43

표 A_1. 1000번의 모의실험을 통한 각 방법의 최적의 기준점 비교	
- Fixed effect(#of 3), Equal variance, Pattern=HH	44
표 A_2. 1000번의 모의실험을 통한 각 방법의 최적의 기준점에서의 Se와 Sp	
- Fixed effect(#of 3), Equal variance, Pattern=HH	45
표 B_1. 1000번의 모의실험을 통한 각 방법의 최적의 기준점 비교	
- Fixed effect(#of 3), Unequal variance, Pattern=HH	46
표 B_2. 1000번의 모의실험을 통한 각 방법의 최적의 기준점에서의 Se와 Sp	
- Fixed effect(#of 3), Unequal variance, Pattern=HH	47
표 C_1. 1000번의 모의실험을 통한 각 방법의 최적의 기준점 비교	
- Random effect(#of 10), Equal variance, Pattern=HH	48
표 C_2. 1000번의 모의실험을 통한 각 방법의 최적의 기준점에서의 Se와 Sp	
- Random effect(#of 10), Equal variance, Pattern=HH	49
표 D_1. 1000번의 모의실험을 통한 각 방법의 최적의 기준점 비교	
- Random effect(#of 10), Unequal variance, Pattern=HH	50
표 D_2. 1000번의 모의실험을 통한 각 방법의 최적의 기준점에서의 Se와 Sp	
- Random effect(#of 10), Unequal variance, Pattern=HH	51
표 E. 1000 번의 모의실험을 통한 두 방법 간 유병률에 따른 최적 기준점과 실제 기준점의 차이 (bias)	
- Random effect(#of 3), Equal variance, AUC=0.961	52
영문요약	53

표 차례

표 1. 진단검사 결과표(1)	5
표 2. 진단검사 결과표(2)	10
표 3. 모의실험 설정표(등분산, 고정효과)	20
표 4. 모의실험 설정표(이분산, 고정효과)	20
표 5. 모의실험 설정표(등분산, 임의효과)	21
표 6. 모의실험 설정표(이분산, 임의효과)	22
표 7. 예제 자료 결과표	30

그림 차례

그림 1. ROC Curve 에서의 Youden index, Euclidean	8
그림 2_1. Fixed effect(# of 3), Equal variance, Pattern=HH일 때, 최적 기준점(AUC=0.702, 0.855, 0.961)	28
그림 2_2. Fixed effect(# of 3), Unequal variance, Pattern=HH일 때, 최적 기준점(AUC=0.702, 0.855, 0.961)	28
그림 3_1. Random effect(# of 10), Equal variance, Pattern=HH일 때, 최적 기준점(AUC=0.702, 0.855, 0.961)	29
그림 3_2. Random effect(# of 10), Unequal variance, Pattern=HH일 때, 최적 기준점(AUC=0.702, 0.855, 0.961)	29

국 문 요 약

여러 평가자들의 진단자료에서 질병 예측의 최적 기준점 선정에 대한 비교 연구

의학 분야에서 질병 진단을 위해 영상 또는 혈액검사 등 여러 진단 도구들을 활용한다. 이러한 진단도구들의 진단 정확도를 평가하는 것은 필수적인 과정이다. 진료 환경에서 유용하게 사용하기 위해 질병 예측력을 최대로 하는 최적의 기준점을 선정하는 과정 또한 필수적이다.

특히 영상 판독 결과를 질병 진단에 사용하는 경우 평가자의 주관이 개입될 가능성이 존재한다. 따라서 본 논문은 여러 평가자 간 변이를 고려하여 종합적인 질병 진단 평가 결과를 이용한다. 종합적인 평가 결과로 진단 도구의 정확성 평가와 최적 기준점 선정은 Multi-Reader Multi-Case (MRMC) ROC Curve를 이용하여 수행한다. MRMC ROC Curve 분석 방법은 모수적 방법 (Gallas, 2006)과 ANOVA를 기반으로 분석하는 비모수적 방법이 있다. 비모수적 방법은 DBM 모형 (Dorfman, Berbaum, and Metz, 1992)과 OR모형 (Obuchowski and Rockette, 2000)이 대표적으로 사용된다.

MRMC ROC Curve의 연구는 활발하게 진행되어 왔지만, MRMC ROC Curve에서 최적의 기준점을 선정하는 방법에 대한 선행 제안 및 비교 연구 사례는 존재하지 않

는다. 따라서 본 연구는 평가자가 한 명일 때, ROC Curve에서 최적의 기준점을 찾는 방법들(Efficiency, Youden index, Euclidean, Likelihood Ratio, Mutual Information)을 MRMC ROC Curve에 적용하여 비교하였다.

본 연구의 모의실험은 여러 상황들을 가정하여 방법에 따라 선정되는 최적 기준점을 살펴보았다. 그 결과, 두 모형(고정효과, 임의효과)의 두 가정(등분산, 이분산)에서 유사한 패턴으로 최적의 기준점이 선정되는 것을 확인할 수 있었다. 방법 Youden index와 Euclidean은 유병률과 관계없이 일정한 최적의 기준점을 선정하는 것을 확인하였다. 그러므로 모집단의 유병률과 표본집단의 유병률의 편차가 심한 경우, 방법 Youden index와 Euclidean을 사용할 것을 권장한다. 이외의 경우는 연구자의 연구목적에 따라 방법을 선정하는 것을 권장한다.

핵심되는 말 : 최적의 기준점, MRMC ROC Curve,
Efficiency, Youden index, Euclidean,
Likelihood Ratio, Mutual Information

제 1장 서론

1.1 연구 배경 및 목적

의료분야에서는 질병 진단을 위해 영상, 혈액검사 등 여러 진단 도구들이 활용된다. 이러한 진단도구의 진단 정확도 평가는 활용 전, 필수적인 과정이다. 진단도구의 정확도를 평가하는 방법은 Receiver-Operating Characteristic (ROC) Curve 아래 면적(AUC; Area Under the ROC Curve)을 측정하여 평가하는 방법이 대표적이다. 또한 질병 유무에 대한 기준점이 알려져 있지 않은 경우, ROC Curve 를 통해 질병 예측력을 최대화시키는 최적의 기준점(threshold)을 찾을 수 있다.

영상 진단 결과는 평가자의 주관이 개입될 수 있어 여러 평가자 간 변이를 고려하기 위해, 여러 평가자의 평가 결과를 종합적으로 사용한다. 이러한 경우, Multi-Reader Multi-Case (MRMC) ROC Curve 를 진단도구의 정확도 평가에 이용한다. MRMC ROC Curve 는 여러 평가자들(readers)이 다양한 진단도구(modalities)를 가지고 여러 대상자들(cases)에 대해서 질병 여부를 예측하는 경우, 예측의 정확성 평가와 예측력을 최대로 하는 최적의 기준점을 찾는 데에 유용하게 사용된다 (Chen, Samuelson, 2014).

MRMC ROC Curve를 분석하는 방법은 크게 두 가지로 나누어진다. 비모수적 방법과(Gallas, 2006) 모수적 방법으로 Analysis of variance (ANOVA)에 기반해서 분석하는 방법이다. 모수적으로 ANOVA에 근거하여 분석하는 방법 중, 가장 보편적으로 사용되는 방법은 Dorfman, Berbaum, and Metz (DBM) 모형을 이용하는 방법이다(Dorfman, Berbaum, Metz, 1992). 평가자들, 대상자들을 임의효과(random

effect)로 가정하고 진단 도구들을 고정효과(fixed effect)로 가정하여 ANOVA 분석법을 적용한다. 대상자들을 하나의 요인으로 가정하기 때문에, 모형이 매우 복잡하여 추정의 한계점들이 발견되었다. 이러한 문제를 보완하기 위해, DBM 모형에서 대상자들을 효과로 고려하지 않은 Obuchowski and Rockette (OR)모형을 이용하는 방법이 제안되었다(Obuchowski and Rockette, 1995). 가장 최근에는 OR모형의 자유도 추정을 보완한 Obuchowski-Rockette-Hillis (ORH) 방법이 제안되었다 (Hillis, 2007).

MRMC ROC Curve 에 대한 활발한 연구에도 불구하고, MRMC ROC Curve 에서 최적의 기준점을 찾는 방법들에 대해 제안 및 비교 연구된 바가 없다. 따라서 본 연구는 평가자가 한 명일 때, ROC Curve 에서 최적의 기준점을 찾는 방법들(Efficiency, Youden index, Euclidean, Likelihood Ratio, Mutual Information)을 MRMC ROC Curve 에 적용하여 위 방법들을 비교하고자 한다.

1.2 연구 내용 및 방법

본 연구의 모의실험에서는 Dorfman, Berbaum 과 Metz 가 제안한 DBM 모형을 수정하여 평가자들의 점수를 생성하는 데에 이용한다. 기존의 DBM 모형은 평가자들, 대상자들을 임의효과로, 진단 도구들을 고정효과로 가정한 모형이고, 수정한 DBM 모형은 평가자들, 대상자들을 임의효과로 가정하고 진단 도구들을 효과로 고려하지 않은 모형이다. 유병률, AUC, 임의효과 분산의 가정들을 설정하여 수정한 OR 모형에 적용한다. 설정에 따라 생성된 점수를 이용하여 각 평가자의 ROC Curve 를 구하고, 이를 통합하여 Pooled ROC Curve 를 구한다.

Pooled ROC Curve 를 이용함으로써, 여러 평가자의 결과를 종합적으로 판단할 수 있는 최적의 기준점을 찾을 수 있다. 이러한 MRMC ROC Curve 에서 최적의 기준점을 찾기 위한 다섯 가지 방법을 제안 및 비교하고자 한다.

1.3 논문의 구성

제 1장에서는 연구의 배경 및 목적, 연구의 내용 및 방법을 소개한다. 제 2장에서는 질병의 예측력을 최대로 하는 최적의 기준점을 구하는 방법들을 살펴본다. 제 3장에서는 모의실험을 통해 MRMC ROC Curve 에서 위 다섯 가지 방법을 적용하여 비교한다. 제 4장에서는 다섯 가지 방법을 실제 자료에 적용하여 결과를 비교한다. 마지막으로 제 5장에서는 결론 및 고찰에 대해 논의한다.

제 2장 이론적 배경

1. Pooled ROC Curve

진단도구를 평가한 평가자 각각의 ROC Curve에서 각 곡선이 변하는 지점을 통합시킨 후, 통일된 각 지점에서의 민감도와 특이도의 평균을 계산한다. 평균 민감도와 특이도를 x축은 1-특이도, y축은 민감도인 좌표평면 위의 곡선으로 표현한 것을 Pooled ROC Curve라고 정의한다. 평균 민감도와 특이도를 수식으로 표현하면 다음과 같다.

진단력 평가에서 $X_j^{(i)}$ 는 i 명의 질병을 가지고 있지 않은 대상자에 대해 j 번 째 평가자가 평가한 점수로 정의한다. 이때, $i = 1, \dots, m$ 이고 $j = 1, \dots, J$ 이다. 이와 같이, $Y_l^{(k)}$ 는 k 명의 질병을 가지고 있는 대상자에 대해 l 번 째 평가자가 평가한 점수로 정의한다. 여기서 $l = 1, \dots, J$ 이고 $k = 1, \dots, n$ 이다.

만약 점수 z 의 높은 값이 질병을 가지고 있는 경우를 의미한다면, 평균 민감도와 특이도는 아래와 같이 정의한다(Lee and Rosner, 2001).

$$\sum_{k=1}^n \sum_{l=1}^J \mathbf{1}(Y_l^{(k)} > z) / nJ$$

$$\sum_{i=1}^m \sum_{j=1}^J \mathbf{1}(X_j^{(i)} < z) / mJ$$

2. 최적 기준점 선정 방법

위 정의된 평균 민감도와 특이도를 최적 기준점을 찾는 데에 이용한다. 그러므로 이 장에서 소개하는 최적 기준점 선정 방법에 사용된 민감도와 특이도는 평균 민감도와 특이도를 의미한다.

2.1 Efficiency (EF)

진단도구의 진단 정확도는 정확하게 분류된 대상자들의 비율로 표현한다. 즉 질병이 있는 대상자들이 진단결과 양성으로 판단되는 경우와 질병이 없는 대상자들이 진단결과 음성으로 판단되는 경우의 합을 전체 대상자 수로 나누어 주는 방법이다. 이를 ‘진단의 효율성’ (diagnostic effectiveness, Efficiency) 또는 ‘정확도’ (accuracy)라고 정의하고, 이 지표의 공식은 다음과 같다[표 1].

표 1. 진단검사 결과표(1)

	질병 유 (D_+)	질병 무 (D_-)
검사결과 양성 (T_+)	a (TP)	c (FP)
검사결과 음성 (T_-)	b (FN)	d (TN)

$$EF = \frac{a + d}{a + b + c + d} \quad or \quad \frac{TP + TN}{TP + FP + FN + TN}$$

이 방법은 질병의 유병률에 의해 영향을 받는다. 질병의 유병률이 감소할수록 검사의 정확성은 증가하는 패턴을 보인다. 위 공식을 질병의 유병률, 민감도와 특이도로 다시 표현하면 아래와 같이 정리할 수 있다. 이 값을 최대로 하는 지점을 EF를 통해 구한 최적의 기준점으로 선정한다.

$$EF = \text{Max} \{P * Se + (1 - P) * Sp\}$$

2.2 W.J.Youden 의 J 통계량(Youden' s index)

W.J.Youden(1950)은 진단도구의 질병 진단 정확도를 평가하기 위해 Youden 지표(index)를 제안하였다. Youden 지표는 ROC Curve 위의 모든 점들 중에서 True Positive Rate (TPR)와 False Positive Rate (FPR)의 차이를 최대로 하는 점을 최적의 기준점으로 설정한다. 이러한 개념을 다음과 같은 두 가지 수식으로 표현할 수 있다(Michils et al, 2009).

$$\begin{aligned} \text{Youden index} &= \text{Max} (\text{TPR} - \text{FPR}) \\ &= \text{Max} \{Se - (1 - Sp)\} \end{aligned}$$

즉 TPR를 최대로, FPR를 최소로 하는 값을 최적의 기준점으로 설정한다는 의미이고, 이는 정확도가 가장 높은 지점을 선택한다는 의미이다. 만약 $F(x)$ 가 ROC Curve의 함수이고, $x=1-\text{특이도}$ 라면, 아래와 같은 함수형태로 정리할 수 있다.

$$J(x) = F(x) - x$$

J 값을 최대로 할 때, $J'(x) = 0$, 다음 수식이 성립한다.

$$J'(x) = F'(x) - 1 = 0$$

J 값이 최대일 때 $F'(x) = 1$ 이다. 이는 J 이 최대인 지점에서 ROC Curve의 탄젠트(tangent), 즉 기울기가 1이 된다는 것을 의미한다. J가 최대인 지점에서의 가로축 좌표는 가로축으로부터 기울기가 1인 직선까지의 거리와 동일하다. 그러므로 Youden 지표는 그림 1과 같이 표현할 수 있다.

2.3 유클리디안(Euclidean)

고대 수학자 Euclid가 제안한 유클리디안 거리(Euclidean distance)를 기본 개념으로 한다. 유클리디안 거리는 다차원 공간에서 두 점 사이의 직선의 거리를 의미한다. 만약 직교 좌표계에 나타낸 점 $p = (p_1, p_2, \dots, p_n)$ 와 $q = (q_1, q_2, \dots, q_n)$ 가 있을 때 두 점 사이의 거리를 구하고자 한다면, 다음 공식을 이용하여 구할 수 있다.

$$d(p, q) = d(q, p) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2} = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

유클리디안(Euclidean) 지표 는 유클리디안 거리의 개념을 ROC Curve에서 최적의 기준점을 구하는 데에 적용한 방법이다. 점 (0,1)과 ROC Curve 위의 어떤 점 사이의 거리를 최소로 하는 기준점을 최적의 기준점으로 설정한다.

$$\text{Euclidean} = \text{Min} \sqrt{(1 - \text{Se})^2 + (1 - \text{Sp})^2}$$

이는 민감도와 특이도를 최대로 하는 최적의 기준점을 설정한다는 의미와 동일하다. Euclidean 지표를 ROC Curve에서 직선으로 표현하면 그림 1과 같다.

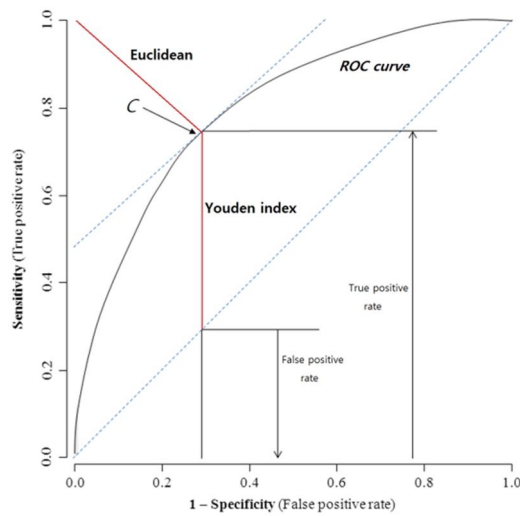


그림 1. ROC Curve에서의 Youden index, Euclidean

2.4 우도비(Likelihood Ratio)

Albert(1982)는 진단검사에서 우도비(Likelihood Ratio) 추정을 제안 하였다. 진단검사의 민감도와 특이도 자료에서 질병의 사후확률 (posterior 또는 post-test probability)을 추정하기 위해 사용하는 우도비의 확장으로 접근한다. 사후확률은 검사 결과를 알고 있을 때 사후현상을 평가하는 데에 사용한다. 그러므로 사후확률을 진단도구의 질병 진단 정확도를 평가하는 데에 사용할 수 있다. 이를 진단검사에 관해서 잘 알려져 있는 이론으로부터 유도하는 과정은 아래와 같다(Johnson, 2004).

진단 정확도는 진단검사 결과 (T) 와 실제 질병 여부 (D) 를 비교하여 산출한다. 검사결과가 양성(T_+)으로 나오는 경우와 음성(T_-)으로 나오는 경우, 실제 질병이 있는 경우 (D_+) 와 없는 경우 (D_-) 에 대해 생각할 수 있다. 민감도는 실제 질병이 있는 대상자들이 진단결과 양성으로 판단되는 확률을 의미하고, 특이도는 실제 질병이 없는 대상자들이 진단결과 음성으로 판단되는 확률을 의미한다. 이를 간단히 수식으로 표현하면 아래와 같다.

$$Se = P(T_+|D_+)$$

$$Sp = P(T_-|D_-)$$

모집단에서 질병의 유병률은 사전확률이라고 부르며 이와 같이 표현한다.

$$P = P(D_+)$$

위의 개념들을 이용하여 다음과 같이 2×2표로 정리할 수 있다[표 2].

표 2. 진단검사 결과표(2)

	질병 유(D_+)	질병 무(D_-)
검사결과 양성(T_+)	PSe	$(1 - P)(1 - Sp)$
검사결과 음성(T_-)	$P(1 - Se)$	$(1 - P)Sp$
Total	P	$1 - P$

진단결과 양성으로 나온 사람들 중에서 실제 질병이 있을 확률을 사후확률이라고 부르며 아래와 같이 표현한다.

$$p_+ = P(D_+|T_+)$$

사후확률은 베이즈 이론(Bayes' theorem)을 적용하여 다음과 같이 표현할 수 있다.

$$P_+ = \frac{P(D_+)P(T_+|D_+)}{P(D_+)P(T_+|D_+) + P(D_-)P(T_+|D_-)} = \frac{PSe}{PSe + (1 - P)(1 - Sp)}$$

사후 오즈(post-odds)는 질병에 걸리지 않을 사후 확률에 대한 질병에 걸릴 사후 확률의 비이다. 식을 이용하여 정리하면 아래 두 식과 같다.

$$\frac{P_+}{1 - P_+} = \frac{Se}{1 - Sp} \frac{P}{1 - P}$$

즉 검사결과 양성으로 판단 된 대상자 중 실제 질병을 가지고 있지 않을 확률에 대한 검사결과 양성으로 판단 된 대상자 중 실제 질병을 가지고 있을 확률의 비이다.

$$\frac{P(D_+|T_+)}{1 - P(D_+|T_+)} = LR_+ * \frac{P(D)}{1 - P(D)}$$

사전확률을 이용한 오즈에 양의 우도비를 곱하여 사후확률을 이용한 오즈를 구한다. 양의 우도비는 실제 질병이 있는 대상자들이 질병이 없는 대상자들에 비해 진단결과가 양성으로 판단되는 확률의 비이다.

$$LR_+ = \text{Max} \left(\frac{Se}{1 - Sp} \right) = \text{Max} \left(\frac{P(T_+|D_+)}{P(T_+|D_-)} \right)$$

이와 같이 사후확률을 이용한 오즈를 구하는 과정에서 유도된 우도비를 ROC Curve에서 최적의 기준점을 구하는 데에 활용한다. 우도비는 ROC Curve에서 기울기를 의미하고, 이를 최대로 하는 지점을 최적의 기준점으로 선정하는 방법이다.

2.5 상호 정보(Mutual Information)

Gilbert Reibnegger(2013)은 진단도구의 질병 진단 정확도를 평가하기 위한 정보 이론의 개념을 처음 제안하였다. 상호 정보의 복잡성과 사용할 수 있는 소프트웨어의 부족 때문에 의학적인 문헌에서 드물게 사용되는 개념이지만, 상당한 이론적 배경을 가지고 있고 2x2표보다 높은 복잡성을 가지고 있는 진단 분석의 상황에 직접적으로

사용할 수 있다는 장점을 가지고 있다.

진단 검사를 수행하기 전, 질병 유무에 대해 의심되는 환자들에 대한 사전확률이 존재한다. 이 때, 환자의 실제 상태에 대해서는 불확실성이 존재한다. 민감도와 특이도를 가진 진단검사를 시행한 후, 우리는 베이즈 이론(Bayes' theorem)을 사용해서 질병 유무의 사후확률을 예측할 수 있다. 검사결과는 위에서 언급한 사전 불확실성을 감소시킨다. 정보 이론에서 환자들은 신호의 원천 (임상적 증상들, 진단검사의 결과들, 등)이 되는 “발송자(sender)”로 간주하고 의사들은 환자들로부터 받은 신호들을 판독하는 역할을 하기 때문에 “수신자(receiver)”로 간주한다. 수신자가 신호(=검사결과)를 받은 후에도 발송자의 실제 상태에 대한 불확실성은 아마도 존재할 것이다. 왜냐하면, 사용되는 진단 검사들은 민감도와 특이도에 영향을 미치는 위양성(false positive) 또는 위음성(false negative)의 판단을 내릴 가능성이 존재하기 때문이다.

정보이론적 개념에서 불확실성은 확률과 관련된다. Shannon (Shannon, 1948)은 모든 가능성 있는 사건들의 각각의 확률들에 대한 불확실성을 정의하기 위해서 로그방법을 활용하였다. 불확실성 정의를 “엔트로피(entropy)”라고 부르고 아래 공식으로 구한다.

$$H(X) = - \sum P(x) \log_2 P(x)$$

여기에서 H는 엔트로피, X는 확률밀도함수 P(x)를 가지고 있는 이분형인 임의의 변수이다 (Reibnegger, 2013).

실제 진단검사는 완벽하지 않기 때문에 민감도와 특이도는 100%보다 작게 추정될 것이다. 그러므로 실제 진단검사를 할 때, 다음과 같이 정보 이론적 처리를 생각할 수 있다. 엔트로피 $H(D)$ 는 진단의 사전확률에 의존하고, 엔트로피 $H(T)$ 는 검사결과의 확률에 의존한다. 만약 진단의 사전확률과 진단결과가 상호적인 관계가 있다면, $H(D)$ 와 $H(T)$ 는 겹쳐질 것이고, 결합 엔트로피 $H(D, T)$ 는 $H(D) + H(T)$ 보다 작을 것이다. 상호적인 정보(Mutual Information)는 이 겹쳐지는 범위로써 정의된다.

$$I(D; T) = H(D) + H(T) - H(D, T)$$

차이 $H(D) - I(D; T)$ 는 엔트로피 $H(D|T)$ 이다. 즉 진단검사를 실시한 후 환자의 실제 질병 상태에 대한 불확실성을 의미한다. 차이 $H(T) - I(D; T)$ 는 엔트로피 $H(T|D)$ 이다. 이는 실제 질병 상태가 주어졌을 때 진단검사결과에 대한 불확실성을 의미한다. 따라서 상호 정보를 증가시킨다는 것은 불확실성을 감소시킨다는 의미이므로, 실제 진단 검사 평가에서 이러한 성질을 유용하게 활용할 수 있다.

상호 정보 지표의 공식 유도 과정은 다음과 같다. 상호 정보는 결합분포 $P(x, y)$ 와 서로 독립적인 두 확률변수 x 와 y 의 곱분포(product 분포) 사이에 상대적인 엔트로피로써 정의할 수 있다(Mackay, 2003).

$$I(X, Y) = \sum \sum p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)}$$

위 개념을 2×2 분류표에 적용하여, 상호 정보를 TPR와 FPR로 표현할 수 있다. 상호 정보 공식을 이용하여 분류표의 확률을 대입하면 아래와 같은 공식을 유도할 수 있다[표 1].

$$I(X, Y) = \frac{a}{n} \log_2 \frac{an}{(a+c)(a+b)} + \frac{b}{n} \log_2 \frac{bn}{(b+d)(a+b)} \\ + \frac{c}{n} \log_2 \frac{cn}{(a+c)(c+d)} + \frac{d}{n} \log_2 \frac{dn}{(b+d)(c+d)}$$

만약, 각각의 경우가 발생할 확률을 알지 못하고, TPR와 FPR만 알고 있다면, 각 경우가 발생할 확률을 TPR(t)와 FPR(f)로 바꾸어 정리한 후, 대입하여 구할 수 있다.

$$a + c = b + d = \frac{n}{2}$$

$$a = \frac{nt}{2}, \quad b = \frac{nf}{2}, \quad c = \frac{n(1-t)}{2}, \quad d = \frac{n(1-f)}{2}$$

이를 상호 정보의 식에 대입하면 아래와 같이 정리할 수 있다.

$$I(X, Y) = \text{Max}[0.5[2 * \log(2) + Se * \log(Se) + (1 - Sp) * \log(1 - Sp) \\ - \{Se + (1 - Sp)\} * \log\{Se + (1 - Sp)\} + (1 - Se) * \log(1 - Se) \\ + Sp * \log(Sp) + \{Sp + (1 - Sp)\} * \log\{Sp + (1 - Sp)\}]]$$

정리된 위의 상호 정보 지표를 최대화하는 지점을 최적의 기준점으로 선정한다.

제 3장 모의실험

제 3장에서는 제 2장에서 소개한 최적의 기준점을 선정하는 다섯 가지 방법에 대해 다음 두 가지 조건을 고려하여 모의실험을 통해 비교하고자 한다.

여러 평가자들이 여러 대상자들에 대해서 질병 여부를 예측하는 연구에서 평가자들의 수는 단일 기관 연구의 한계로 인해 흔히 두 명 또는 세 명으로 설정한다. 하지만 평가자들이 많을수록 예측 결과의 모집단 대표성이 높아지므로 여러 평가자가 참여한 연구를 하는 경우 적어도 네 명 이상의 평가자와 연구하는 것을 권장한다 (Berbaum et al, 2009). 만약 제한사항으로 인해 평가자를 네 명 미만으로 설정해야 한다면 평가자를 고정효과로 고려하여 자료 분석을 진행한다. 따라서 본 모의실험은 평가자들의 수를 고려하여 평가자를 고정효과로 가정한 모형과 임의효과로 가정한 모형으로 나누어 시행한다.

이론적으로 실제 질병이 있는 집단과 실제 질병이 없는 집단의 진단 결과의 분산이 동일한 경우를 가정한다. 하지만, 실제 자료에서는 실제 질병이 있는 사람들의 평가점수의 분산과 실제 질병이 없는 사람들의 평가점수의 분산이 항상 동일하지 않기 때문에 위 두 가지 모형 하에서 분산이 동일한 경우와 동일하지 않은 경우를 고려하여 모의실험을 시행한다.

각 모의실험에서 평가 점수는 질병의 유병률, AUC, 임의효과의 분산에 따라 생성한다.

3.1 모의실험 설계

선형복합모델(LMM; Linear Mixed Model)을 통해, 평가자의 점수(X_{ijt})를 임의로 생성한다. 모형은 Dorfman et al(1998) 논문의 제안된 model을 수정한 모형이다.

$$X_{ijt} = \mu_t + R_{it} + C_{jt} + (RC)_{ijt} + \varepsilon_{ijt}$$

위의 모형의 의미는 다음과 같다.

i =평가자의 수, j =질병 발생자 수, t =실제 질병 여부(0: 질병 무, 1: 질병 유),
 μ_t =평가자&대상자의 절편(0: $t=0$, a/b: $t=1$), R_{it} =평가자, C_{jt} =대상자,
 $(RC)_{ijt}$ =평가자*대상자 ε_{ijt} =오차를 의미한다.

평가자를 고정효과로 고려하는 경우 평가자의 수는 3명($i=3$)으로 설정하고,
 평가자를 임의효과로 고려하는 경우 평가자의 수는 3명과 10명($i=3, 10$)으로
 설정한다. 질병 발생자 수는 전체 N수(N=100,200)와 질병의 유병률(P=10%, 30%,
 50%, 80%)에 따라 다르게 설정한다.

평가자를 고정효과로 가정한 모형에서 요인 μ_t 와 R_{it} 는 고정효과로, 나머지 요인
 C_{jt} , $(RC)_{ijt}$, ε_{ijt} 들은 임의효과로 설정한다. 실제 질병이 있는 집단과 실제 질병이
 없는 집단의 진단 결과의 분산의 동일성 여부에 따라 고정효과와 임의효과를
 설정한다. 고정효과는 (i) 실제 질병이 있는 집단과 질병이 없는 집단의 분산이
 동일한 경우, (ii) 실제 질병이 있는 집단과 질병이 없는 집단의 분산이 동일하지
 않은 경우의 설정으로 아래와 같다. 임의효과는 평균이 0이고 분산은 위의 설정을
 따르는(표 3, 표 4) 정규분포에서 생성한다.

$$(i) \mu_0 = 0, \mu_1 = 0.75, 1.5, 2.5$$

$$(ii) \mu_0 = 0, \mu_1 = 0.821, 1.831, 3.661$$

$$R_{1t} = -1, R_{2t} = 0, R_{3t} = 1$$

μ_t 는 AUC값에 따라 설정하고, R_{it} 는 평가자의 점수를 생성하는 데에 큰 영향을 주지 않는 범위 안에서 위와 같이 설정한다.

평가자를 임의효과로 가정한 모형에서 유일하게 요인 μ_t 를 고정효과로 설정한다. 고정효과 μ_t 는 평가자가 고정효과인 모형의 μ_t 설정 값과 동일하게 설정하고, 임의효과는 다음의 평균과 분산을 따르는 정규분포에서 생성한다. 임의효과의 평균은 0으로 설정하고 분산은 실제 질병이 있는 집단과 실제 질병이 없는 집단의 분산의 동일성 여부에 따라 설정한다(표 5, 표 6).

평가자를 고정효과로 가정한 모형과 임의효과로 가정한 모형에서 각각 실제 질병이 있는 집단과 질병이 없는 집단의 분산이 동일한 경우와 동일하지 않은 경우로 나누었을 때, AUC의 값과 상관관계의 값은 동일하게 설정한다. 상관관계는 두 가지로 정의하는데, 하나는 평가자들 간의 평가점수의 상관관계를 의미하고 다른 하나는 평가자 내의 평가점수의 상관관계를 의미한다.

AUC값의 수식은 아래와 같다.

$$AUC = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right)$$

여기서 a는 실제 질병이 있는 사람들의 평균에서 실제 질병이 없는 사람들의 평균을

뺀 값을 의미하고, b는 실제 질병이 있는 사람들의 표준편차에 대한 실제 질병이 없는 사람들의 표준편차의 비를 의미한다. AUC값은 a와 b의 설정을 통해 두 경우에서 모두 0.702, 0.855, 0.962로 동일하게 설정한다(Dorfman, 1998).

상관관계는 임의효과의 분산들을 이용하여 정의할 수 있다. 평가자 내의 평가 점수의 상관성(ρ_{WR})과 평가자들 간의 평가 점수의 상관성(ρ_{BR})을 구하는 공식은 다음과 같다.

$$\rho_{WR} = (\sigma_C^2 + \sigma_{RC}^2) / (\sigma_C^2 + \sigma_{RC}^2 + \sigma_\epsilon^2)$$

$$\rho_{BR} = (\sigma_C^2) / (\sigma_C^2 + \sigma_{RC}^2 + \sigma_\epsilon^2)$$

$\rho_{WR}=0.8$, $\rho_{BR}=0.6$ 인 경우 상관관계가 높다고 해석하고, $\rho_{WR}=0.4$, $\rho_{BR}=0.2$ 인 경우 상관관계가 낮다고 해석한다. 분산의 패턴(pattern)에 따라서 상관관계의 높고 낮음이 설정되고 이를 H (High), L (Low)로 표현한다.

평가자를 임의효과로 가정한 모형에서 실제 질병이 있는 집단과 질병이 없는 집단의 분산이 동일한 경우와 동일하지 않은 경우의 설정은 다음과 같은 차이점을 가지고 있다. 실제 질병의 유무 집단의 분산이 동일하다고 가정할 때, 임의효과로 가정한 평가자의 분산(σ_R^2)이 0.011, 0.030, 0.056인 경우 분산이 크다고 해석하고, 0.005인 경우 분산이 작다고 해석한다. 실제 질병의 유무 집단의 분산이 동일하지 않다고 가정할 때, 임의효과로 가정한 평가자의 분산(σ_R^2)이 0.0132, 0.0447, 0.1201인 경우 분산이 크다고 해석하고, 0.0066, 0.0082, 0.0118인 경우 분산이 작다고 해석한다. 상관관계의 표현법과 동일하게 분산의 크고 작음을 H, L로 표현한다. 이 두 가지의 패턴을 하나로 결합하여 표현하는데, 첫 번째 문자는 상관관계의 정도를 의미하고 두 번째 문자는 분산의 크기를 의미한다. 이들의 조합은 HL, LL, HH, LH로 정리할 수 있다(표 5, 표6).

표 3. 모의실험 설정표(등분산, 고정효과)

Pattern	μ_+	AUC	Correlation		Variance components		
			ρ_{WR}	ρ_{BR}	σ_C^2	σ_{RC}^2	σ_ε^2
H	0.75	0.702	0.8	0.6	0.6	0.2	0.2
H	1.50	0.855	0.8	0.6	0.6	0.2	0.2
H	2.50	0.962	0.8	0.6	0.6	0.2	0.2
L	0.75	0.702	0.4	0.2	0.2	0.2	0.6
L	1.50	0.855	0.4	0.2	0.2	0.2	0.6
L	2.50	0.962	0.4	0.2	0.2	0.2	0.6

표 4. 모의실험 설정표(이분산, 고정효과)

Pattern	μ_+	AUC	Correlation		Variance Components					
			ρ_{WR}	ρ_{BR}	$\sigma_{C(-)}^2$	$\sigma_{RC(-)}^2$	$\sigma_{\varepsilon(-)}^2$	$\sigma_{C(+)}^2$	$\sigma_{RC(+)}^2$	$\sigma_{\varepsilon(+)}^2$
H	0.821	0.702	0.8	0.6	0.6	0.2	0.2	0.84	0.28	0.28
H	1.831	0.855	0.8	0.6	0.6	0.2	0.2	1.18	0.40	0.40
H	3.661	0.962	0.8	0.6	0.6	0.2	0.2	1.98	0.66	0.66
L	0.821	0.702	0.4	0.2	0.2	0.2	0.6	0.28	0.28	0.84
L	1.831	0.855	0.4	0.2	0.2	0.2	0.6	0.40	0.40	1.19
L	3.661	0.962	0.4	0.2	0.2	0.2	0.6	0.66	0.66	1.97

표 5. 모의실험 설정표(등분산, 임의효과)

Pattern	μ_+	AUC	Correlation		Variance components			
			ρ_{WR}	ρ_{BR}	σ_C^2	σ_{RC}^2	σ_ε^2	σ_R^2
HL	0.75	0.702	0.8	0.6	0.6	0.2	0.2	0.0055
HL	1.50	0.855	0.8	0.6	0.6	0.2	0.2	0.0055
HL	2.50	0.962	0.8	0.6	0.6	0.2	0.2	0.0055
LL	0.75	0.702	0.4	0.2	0.2	0.2	0.6	0.0055
LL	1.50	0.855	0.4	0.2	0.2	0.2	0.6	0.0055
LL	2.50	0.962	0.4	0.2	0.2	0.2	0.6	0.0055
HH	0.75	0.702	0.8	0.6	0.6	0.2	0.2	0.011
HH	1.50	0.855	0.8	0.6	0.6	0.2	0.2	0.030
HH	2.50	0.962	0.8	0.6	0.6	0.2	0.2	0.056
LH	0.75	0.702	0.4	0.2	0.2	0.2	0.6	0.011
LH	1.50	0.855	0.4	0.2	0.2	0.2	0.6	0.030
LH	2.50	0.962	0.4	0.2	0.2	0.2	0.6	0.056

표 6. 모의실험 설정표(이분산, 임의효과)

Pattern	μ_+	AUC	Correlation		Variance Components						
			ρ_{WR}	ρ_{BR}	$\sigma_{C(-)}^2$	$\sigma_{RC(-)}^2$	$\sigma_{\epsilon(-)}^2$	$\sigma_{C(+)}^2$	$\sigma_{RC(+)}^2$	$\sigma_{\epsilon(+)}^2$	σ_R^2
HL	0.821	0.702	0.8	0.6	0.6	0.2	0.2	0.84	0.28	0.28	0.0066
HL	1.831	0.855	0.8	0.6	0.6	0.2	0.2	1.18	0.40	0.40	0.0082
HL	3.661	0.962	0.8	0.6	0.6	0.2	0.2	1.98	0.66	0.66	0.0118
LL	0.821	0.702	0.4	0.2	0.2	0.2	0.6	0.28	0.28	0.84	0.0066
LL	1.831	0.855	0.4	0.2	0.2	0.2	0.6	0.40	0.40	1.19	0.0082
LL	3.661	0.962	0.4	0.2	0.2	0.2	0.6	0.66	0.66	1.97	0.0118
HH	0.821	0.702	0.8	0.6	0.6	0.2	0.2	0.84	0.28	0.28	0.0132
HH	1.831	0.855	0.8	0.6	0.6	0.2	0.2	1.18	0.40	0.40	0.0447
HH	3.661	0.962	0.8	0.6	0.6	0.2	0.2	1.98	0.66	0.66	0.1201
LH	0.821	0.702	0.4	0.2	0.2	0.2	0.6	0.28	0.28	0.84	0.0132
LH	1.831	0.855	0.4	0.2	0.2	0.2	0.6	0.40	0.40	1.19	0.0447
LH	3.661	0.962	0.4	0.2	0.2	0.2	0.6	0.66	0.66	1.97	0.1201

위 내용을 정리하면 다음과 같다.

각 모의실험은 설정된 AUC 값(=0.702, 0.855, 0.961)에 따라서 분산의 패턴(=HL, LL, HH, LH)과 질병의 유병률(=10%, 30%, 50%, 80%)을 다르게 설정하여 자료를 생성한다. 이와 같은 과정을 1000번 반복하여 진행한다.

각 설정에 따라 생성된 자료에서 Pooled ROC Curve를 구한 후, 다섯 가지 방법(Efficiency, Youden index, Euclidean, Likelihood Ratio, Mutual Information)을 적용하여 비교한다. 모든 분석은 3.2.2버전의 R 프로그램(R Foundation for Statistical Computing, Vienna, Austria)을 사용하여 분석하였으며, 다섯 가지 방법의 공식을 프로그래밍하여 각 방법에 따른 최적의 기준점, 민감도와 특이도를 구하였다. 각 설정에서 방법 마다 1000개의 최적의 기준점이 계산되는데, 이들의 평균을 최종 값으로 한다. 최종적으로 구한 최적의 기준점과 실제 기준점(true cutoff)과의 편차를 상대편차(relative bias)로 정의하여 비교한다.

$$\%Bias(\hat{c}) = \frac{\hat{c} - c}{c} * 100\%$$

여기에서 \hat{c} 은 최적의 기준점을 의미하고, c 는 실제 기준점을 의미한다. 즉 최적의 기준점과 실제 기준점의 차이를 실제 기준점으로 나눈 값에 100%를 곱한 값으로 상대편차의 식을 정의한다.

종합적으로 본 연구의 모의실험을 통해 여러 조건 하에서 유병률에 따른 방법 별 최적의 (평균)기준점을 비교 평가하고자 한다.

3.2 모의실험 결과

모의실험의 결과는 전반적으로 유사한 패턴을 보인다. 모의실험 결과의 전반적인 패턴을 살펴보면 다음과 같다.

그림 A_1, A_2, A_3, A_4는 네 가지 분산의 Pattern 에서 유병률에 따른 방법 별 최적의 기준점의 평균을 나타낸다. 분산의 패턴(pattern)은 평가자 안에서의 평가 점수의 상관성, 평가자들 간의 평가 점수의 상관성과 평가자들의 분산을 고려한다. 그림 A_1, A_2, A_3, A_4의 결과 패턴이 동일한 것을 통해 분산의 패턴은 최적의 기준점을 선정하는 데에 영향을 주지 않는다는 것을 확인할 수 있다. 즉, 평가자 안에서의 평가 점수의 상관성, 평가자들 간의 평가 점수의 상관성과 평가자들의 분산에 관계없이 최적의 기준점이 선정된다. 그러므로 그림 A_1에 대해 대표적으로 살펴보고자 한다.

그림 A_1의 세 가지 결과는 AUC=0.702, 0.855, 0.961에서 유병률에 따른 방법 별 최적 기준점의 평균을 나타낸다. 전체적으로 모든 방법에서 AUC 가 증가할수록 높은 최적의 기준점을 선정하고, 유병률에 영향을 덜 받는 것을 확인할 수 있다.

방법 LR (Likelihood Ratio), MI (Mutual Information), EF (Efficiency)은 유병률에 따라 변화하는데, 유병률이 증가함에 따라 낮은 기준점을 최적의 기준점으로 선정하는 경향이 있다. 특히, 방법 EF 는 다른 두 방법들에 비해 유병률에 영향을 더 많이 받는다. 이에 반해, 방법 Youden index, Euclidean 으로 구한 최적의 기준점은 유병률이 변하여도 거의 일정한 결과를 제시한다.

모의실험 결과를 설정한 두 모형(평가자를 고정효과로 가정한 모형, 평가자를 임의효과로 가정한 모형)에 따라 살펴보면 다음과 같다.

방법 LR 과 MI 은 고정효과로 가정한 모형의 경우, 임의효과로 가정한 모형보다 더 높은 최적의 기준점을 선정한다. 방법 Youden index, Euclidean 은 고정효과로 가정한 모형과 임의효과로 가정한 모형에서 유사한 최적의 기준점을 선정한다. 방법 EF 는 AUC 값이 0.702이고 유병률이 0.8일 때, 고정효과로 가정한 모형에서 구한 최적의 기준점이 임의효과로 가정한 모형에서의 최적의 기준점보다 현저히 낮게 설정된다.

평가자를 고정효과로 가정한 모형보다 임의효과로 가정한 모형의 경우와 실제 질병이 있는 집단과 질병이 없는 집단의 분산이 동일한 경우보다 동일하지 않은 경우에서 각 방법 별 선정되는 최적의 기준점이 유사하다.

3.2.1 평가자를 고정효과로 가정한 모형

평가자를 고정효과로 가정한 모형에서 평가자의 수는 3명으로 설정한 결과이다(그림 1_1, 그림 1_2).

1) 두 집단의 분산이 등분산인 경우

실제 질병이 있는 집단과 실제 질병이 없는 집단의 분산이 동일한 두 정규분포를 따를 때, 방법 Youden index 와 Euclidean 에서 질병을 예측하는 실제 기준점(true cutoff)은 두 집단의 평가 점수의 평균들의 평균 값으로 구한다. AUC=0.702일 때, $\mu_0=0$ $\mu_1=0.75$ 이다. 이는 질병이 없는 집단의 평가 점수의 평균은 0이고, 질병이 있는 집단의 평가 점수의 평균은 0.75라는 의미이다. 따라서 실제 기준점은 이 둘의 평가 점수의 평균인 0.375이다. 그림 A_1의 첫 번째 그림을 보면, 최적 기준점이 0.375 근처에서 안정적으로 제시되는 것을 볼 수 있다. AUC=0.855일 때, $\mu_0=0$ $\mu_1=1.5$ 이고 실제 기준점은 두 평균 평가점수의 평균인 0.75이다. 그림 A_1의 두 번째 그림을 통해, 방법 Youden index, Euclidean 이 0.75 근처에서 최적의 기준점을 제시하는 것을 확인할 수 있다. AUC=0.961일 때, $\mu_0=0$ $\mu_1=2.5$ 로 실제 기준점은 1.25이다. 그림 A_1의 마지막 그림을 통해 방법 Youden index, Euclidean 이 실제 기준점과 유사한 지점에서 최적의 기준점을 선정하는 것을 확인할 수 있다.

위 조건하에서 구한 실제 기준점을 그림과 표로 표현한 결과는 다음과 같다. 그림 B 를 통해 두 방법의 최적의 기준점은 유병률이 증가함에 따라 실제 기준점에 가까워진다는 것을 확인 할 수 있고, 표 A_1, 표 B_1, 표 C_1, 표 D_1, 표 E 를 통해 동일한 결과를 확인할 수 있다. 표 E 는 유병률이 증가함에 따라 %Bias 가 감소하는 패턴을 통해 확인이 가능하다.

2) 두 집단의 분산이 이분산인 경우

두 집단의 분산이 등분산인 경우의 그래프와 비교하면 다음의 특징을 지니고 있다. 최적 기준점의 전체적인 패턴은 유사하나 방법 Youden index 와 Euclidean 으로 구한 최적의 기준점의 경우 다른 양상을 보인다. 두 집단의 분산이 등분산인 경우보다 이분산인 경우 두 방법(Youden index, Euclidean)으로 구한 최적의 기준점의 차이가 크다.

3.2.2 평가자를 임의효과로 가정한 모형

평가자를 임의효과로 가정한 모형에서 평가자의 수는 10명으로 설정한 결과이다(그림 2_1, 그림 2_2). 추가적으로 위 모형에서 평가자의 수를 3명으로 설정하여 모의실험을 실시하였다(그림 C_1, 그림 C_2).

1) 두 집단의 분산이 동일한 경우

두 집단의 분산이 이분산인 경우와 최적의 기준점이 선정되는 전반적인 패턴은 동일하다. 세부적인 차이는 평가자의 수에 따라 살펴보았다.

평가자의 수가 10명일 때, 두 집단의 분산이 등분산인 경우 이분산인 경우보다 높은 최적의 기준점을 선정하는 경향을 보인다.

이와 반대로 평가자의 수가 3명일 때, 두 집단의 분산이 등분산인 경우 이분산인 경우보다 낮은 최적의 기준점을 선정하는 경향을 보인다.

2) 두 집단의 분산이 이분산인 경우

각 방법 별로 최적의 기준점을 선정할 때, 두 집단의 분산이 이분산인 경우 등분산인 경우보다 기준점들이 더 유사하게 선정되는 특징을 가지고 있다. 이러한 특징은 두 가지 모형(평가자를 고정효과로 가정한 모형, 평가자들을 임의효과로 가정한 모형)과 두 가지 경우(실제 질병이 있는 집단과 질병이 없는 집단의 분산이 동일한 경우, 동일하지 않은 경우)에서의 모의실험 결과를 종합적으로 보았을 때, 더 뚜렷하게 볼 수 있다. 이러한 경향은 평가자의 수의 관계없이 동일하다.

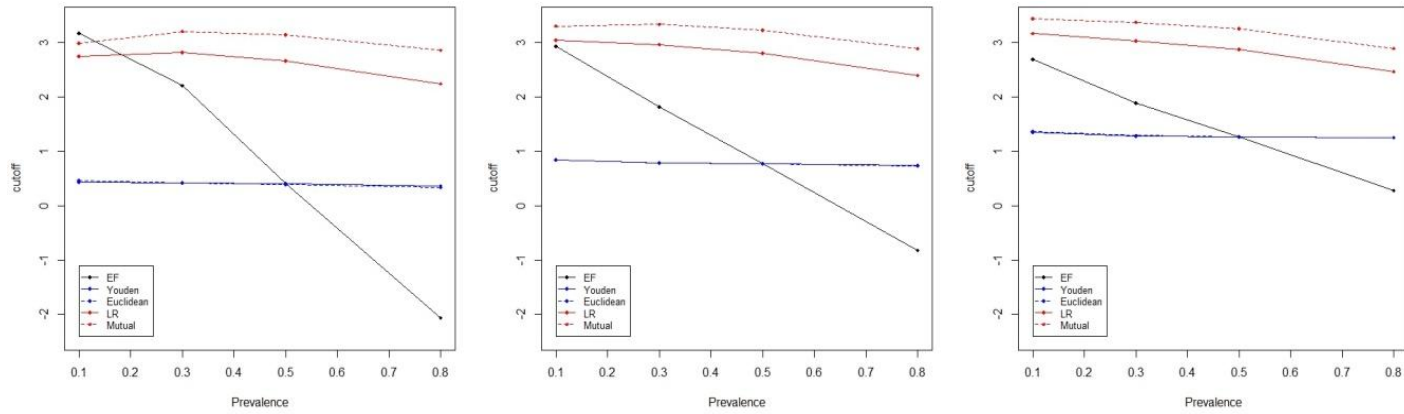


그림 1_1. Fixed effect (# of 3), Equal variance, Pattern=HH일 때, 최적 기준점 (AUC=0.702, 0.855, 0.961)

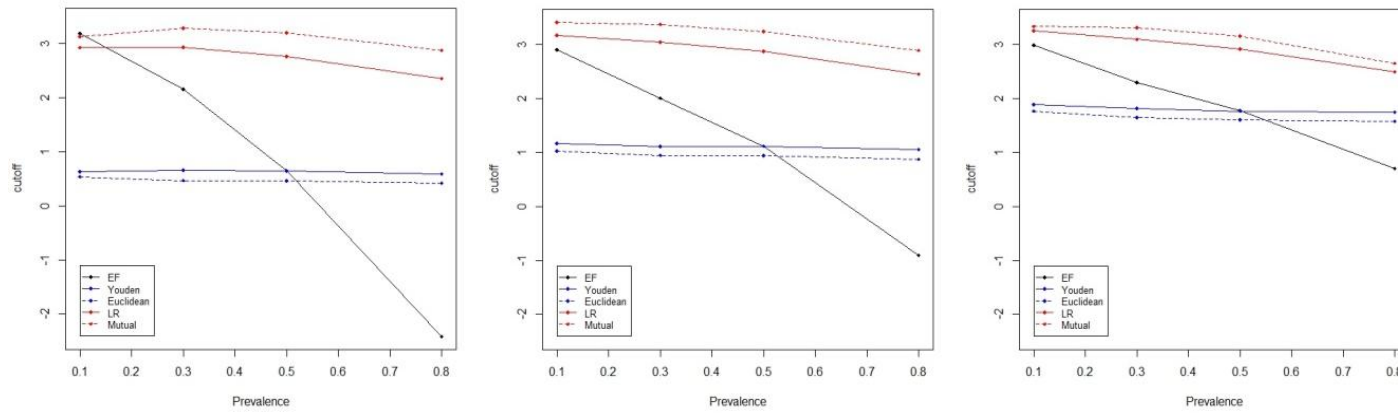


그림 1_2. Fixed effect (# of 3), Unequal variance, Pattern=HH일 때, 최적 기준점 (AUC=0.702, 0.855, 0.961)

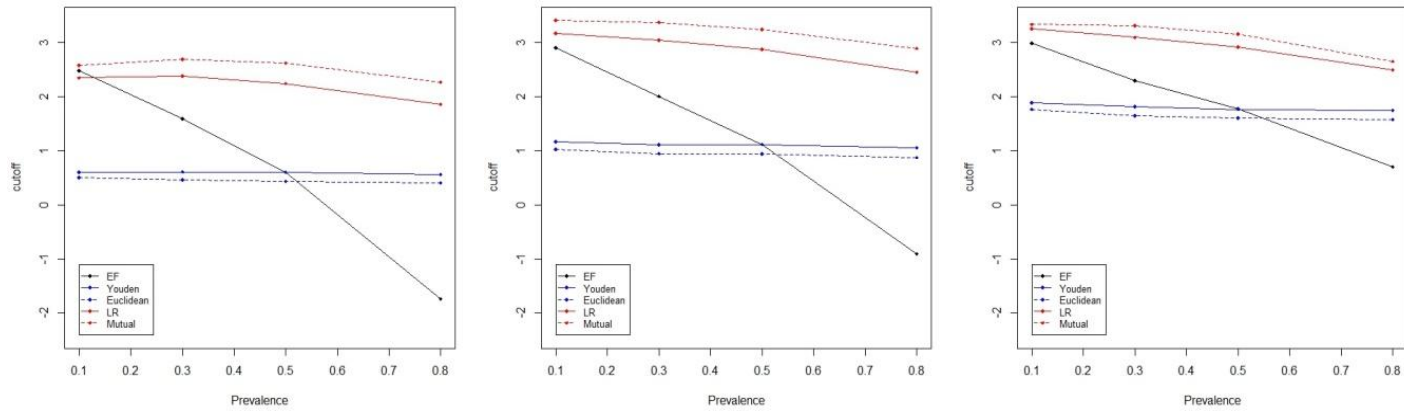


그림 2_1. Random effect (# of 10), Equal variance, Pattern=HH일 때, 최적 기준점 (AUC=0.702, 0.855, 0.961)

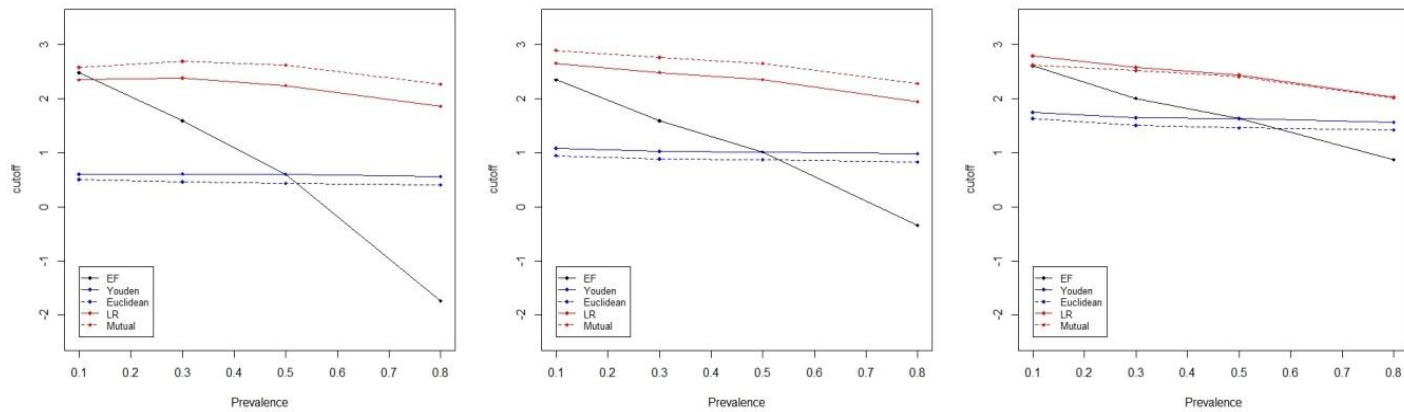


그림 2_2. Random effect(# of 10), Unequal variance, Pattern=HH일 때, 최적 기준점 (AUC=0.702, 0.855, 0.961)

제 4장 실제 자료 적용

4장에서는 실제 임상에서 사용된 자료를 이용하여 다섯 가지 통계량을 비교하고자 한다.

실제 자료는 연세대학교 의과대학 세브란스병원 영상의학과에서 2016년 7월에 완료된 전립선암 진단을 위한 자기 공명(MR; Magnetic Resonance) 영상 평가자료이다. 위 자료는 두 명의 영상의학과 의사에 의해 평가되었다. 암 진단을 위해 MR 영상을 촬영한 환자를 대상으로 32명의 자료를 수집하였다. 평가점수는 5점 척도(1: 매우낮음, 2: 낮음, 3: 보통, 4: 높음, 5: 매우높음)이고, 점수가 높을수록 전립선암에 걸릴 확률이 높다. 자료의 유병률은 0.82이다.

두 평가자의 ROC Curve 를 Pooling 시킨 후, 최적의 기준점을 찾기 위해 다섯 가지 통계량을 적용한 결과는 다음과 같다. Efficiency, Youden index, Euclidean, Mutual Information 으로 구한 최적의 기준점은 모두 3점이고, Likelihood Ratio 로 구한 최적의 기준점은 4점이다. 최적의 기준점이 3점인 경우 민감도는 0.8704이고 특이도는 0.8000이다. 최적의 기준점이 4점인 경우 민감도는 0.3889, 특이도는 1.0000이다(표 7).

표 7. 예제 자료 결과표

Method	Cut-off	Se	Sp
Efficiency	3	0.8704	0.8000
Youden index	3	0.8704	0.8000
Euclidean	3	0.8704	0.8000
Likelihood Ratio	4	0.3889	1.0000
Mutual Information	3	0.8704	0.8000

민감도와 특이도 모두 높은 값을 가진 경우, 평가로 질병을 잘 분류한다고 할 수 있다. 하지만, 예제에서 Likelihood Ratio 로 구한 최적의 기준점으로 질병을 평가하는 경우 특이도는 매우 높지만 민감도는 확연히 떨어지는 것을 확인할 수 있다. 그러므로 Likelihood Ratio 를 이용하여 구한 최적의 기준점 4보다 다른 네 가지 방법으로 구한 최적의 기준점 3으로 질병을 평가하는 것이 질병을 분류하기에 더 적절한 결과라고 판단된다.

제 5장 결론 및 고찰

본 연구는 여러 평가자들이 대상자의 질병 유무를 평가할 때, 질병 유무에 대한 기준점이 알려져 있지 않은 경우 MRMC ROC Curve 를 통해 질병 예측력을 최대화시키는 최적의 기준점을 찾는 방법에 대해 제안한 연구이다. 기존에 제안된 최적의 기준점을 찾는 방법들은 평가자가 한 명 일 때의 상황만을 가정해왔다. 현재, 평가자가 여러 명 인 경우 최적의 기준점을 선정하는 방법에 대한 선행 제안 및 비교 연구 사례가 없다. 따라서 본 연구는 MRMC ROC Curve 에서 단일 ROC Curve 의 최적 기준점을 찾는 방법들(Efficiency, Youden index, Euclidean, Likelihood Ratio, Mutual Information)을 적용하여 평가자가 여러 명 인 경우 최적의 기준점을 선정하는 방법들을 제안 및 비교하였다.

모의실험은 평가자들을 고정효과로 가정한 모형과 임의효과로 가정한 모형에서 실제 질병이 있는 집단과 질병이 없는 집단의 분산이 등분산인 경우와 이분산인 경우로 나누어 실시하였다. 모의실험 결과 두 모형(고정효과, 임의효과)의 두 가정(등분산, 이분산)에서 유사한 패턴으로 최적의 기준점이 선정되는 것을 확인할 수 있었다. 다만 평가자들을 고정효과로 가정한 모형보다 임의효과로 가정한 모형에서 각 방법으로 구한 최적의 기준점들이 유사하다는 것을 확인할 수 있었다. 또한 실제 질병이 있는 집단과 질병이 없는 집단의 분산이 동일한 경우보다 동일하지 않은 경우에서 각 방법으로 구한 최적의 기준점들이 유사하다는 것을 확인할 수 있었다.

평가자들을 고정효과로 가정하고, 실제 질병이 있는 집단과 없는 집단의 분산이 동일하다고 설정한 경우에 Youden index 와 Euclidean 의 실제 기준점을 구할 수

있었다. 하지만 이외의 경우에는 실제 기준점을 구하지 못하기 때문에 특정한 방법으로 최적의 기준점을 구하는 것이 예측력이 높다고 평가할 수 없다.

그러므로 연구자의 연구 목적에 따라 여러 방법 중 가장 적합한 방법으로 최적 기준점을 선정하는 것을 권장한다. 방법 별 최적 기준점은 다음과 같다. 첫째, 방법 Efficiency 의 최적 기준점은 질병 진단 결과 정확하게 분류된 대상자들의 비율을 최대로 하는 지점이다. 이는 유병률과 상관관계가 높다. 둘째, 방법 Youden index 의 최적 기준점은 TPR 와 FPR 의 차이를 최대로 하는 지점이다. 이는 유병률과 상관관계가 거의 존재하지 않는다. 셋째, 방법 Euclidean 의 최적 기준점은 Pooled ROC Curve 의 그래프 상에서의 점 (0,1)과 Pooled ROC Curve 위의 어떤 점 사이의 거리를 최소로 하는 지점이다. 이 또한 유병률과 상관관계가 거의 존재하지 않는다. 넷째, 방법 Positive Likelihood Ratio 의 최적 기준점은 실제 질병이 있는 대상자들이 질병이 없는 대상자들에 비해 진단결과가 양성으로 판단되는 확률의 비를 최대로 하는 지점이다. 마지막으로 방법 Mutual Information 의 최적 기준점은 실제 질병 예측의 불확실성을 최소화하는 지점이다.

본 연구의 모의실험 결과를 통해 방법 Youden index 와 Euclidean 은 유병률과 관계없이 일정하게 최적 기준점이 선정되는 것을 확인하였다. 따라서, 모집단의 유병률과 표본집단의 유병률의 편차가 심한 경우, 방법 Youden index 또는 Euclidean 을 최적의 기준점 선정 방법으로 사용하는 것을 권장한다.

실제 자료는 전립선암 진단을 위한 MR 영상 평가자료를 이용하여 분석하였다. 평가점수는 이산형 자료인 5점 척도로 평가하였다. MRMC ROC Curve 에서 다섯 가지 방법을 적용하여 최적의 기준점을 구한 결과 3 또는 4로 모두 유사한 결과를 기준점으로 제시하였다. 모의실험에서 연속형 평가점수를 이용하여 최적의 기준점을

선정하였을 때보다 예제 자료에서 이산형 평가점수를 이용하여 최적의 기준점을 선정하였을 때, 최적의 기준점을 구하는 방법에 관계없이 유사한 결과를 도출하였다.

본 연구에서 이용한 다섯 가지의 최적 기준점 선정 방법 이외에도 다른 방법들이 존재한다. 최적의 기준점을 선정하는 방법 별로 다른 기준을 가지고 선정하기 때문에, 연구 목적에 따라 적합한 최적 기준점 선정 방법이 다를 것이다. 때문에 향후 더 많은 연구가 필요할 것으로 사료된다.

참고문헌

Albert, A. (1982). On the use and computation of likelihood ratio in clinical chemistry. *Clin Chem*, 28:1113–1119.

Berbaum, K. S., et al. (2009). DBM MRMC 2.1 User' s Guide (software manual). <http://perception.radiology.uiowa.edu> (Accessed August 1, 2016).

Chen, W., Samuelson F. W. (2014). The average receiver operating characteristic curve in multireader multicase imaging studies. *Br J Radiol*, 87:20140016.

Dorfman, D. D., Berbaum, K. S., Metz, C. E. (1992). Receiver operating characteristic rating analysis: Generalization to the population of readers and patients with the jackknife method. *Investig Radiol*, 27:723–731.

Dorfman, D. D et al. (1998). Monte Carlo validation of a multireader method for receiver operating characteristic discrete rating data: Factorial experimental design. *Acad Radiol*, 5:591–602.

Gallas, B. D. (2006). One-Shot Estimate of MRMC Variance: AUC. *Acad Radiol*, 13:353–362.

Hillis, S. L. (2007). A comparison of denominator degrees of freedom methods for multiple observer ROC analysis. *Stat Med*, 26:596–619.

Johnson, N. P. (2004). Advantages to transforming the receiver operating characteristic (ROC) curve into likelihood ratio co-ordinates. *Stat Med*, 23:2257–2266.

Lee, M. L. T. and Rosner, B. A. (2001). The average area under correlated receiver operating characteristic curves: a nonparametric approach based on generalized two-sample Wilcoxon statistics. *Appl. Statist*, 50(3):337–344.

Mackay, D. J. C. (2003). Information Theory, Inference, and Learning Algorithms. *Cambridge: Cambridge University Press*, 140:532–533.

Michils, A et al. (2009). Exhaled nitric oxide as a marker of asthma control in smoking patients. *Eur Respir J*, 33:1295–1301.

Obuchowski, N. A., Rockette, H. E. (1995). Hypothesis testing of diagnostic accuracy for multiple readers and multiple tests: An ANOVA approach with dependent observations. *Commun Stat Simul Comput*, 24:285–308.

Reibnegger, G. (2013). Beyond the 2×2 -contingency table: A primer on entropies and mutual information in various scenario involving m diagnostic categories and n categories of diagnostic tests. *Clin Chim Acta*, 423:97–103.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst Technol J*, 27:379-423, 623-656.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3(1):32-35.

부 록

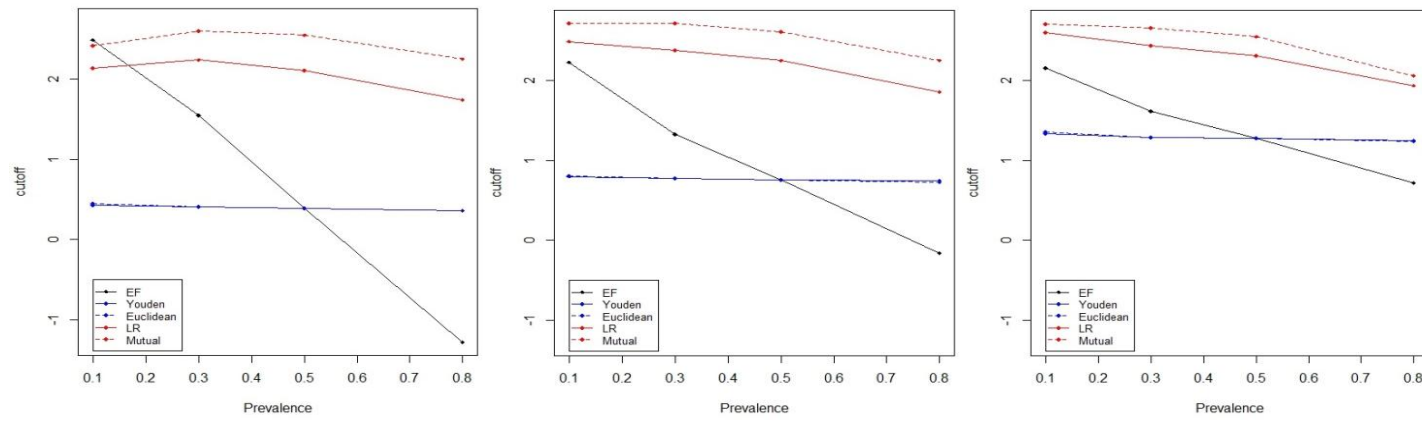


그림 A_1. Random effect (# of 3), Equal variance, Pattern=HL일 때, 최적 기준점 (AUC=0.702, 0.855, 0.961)

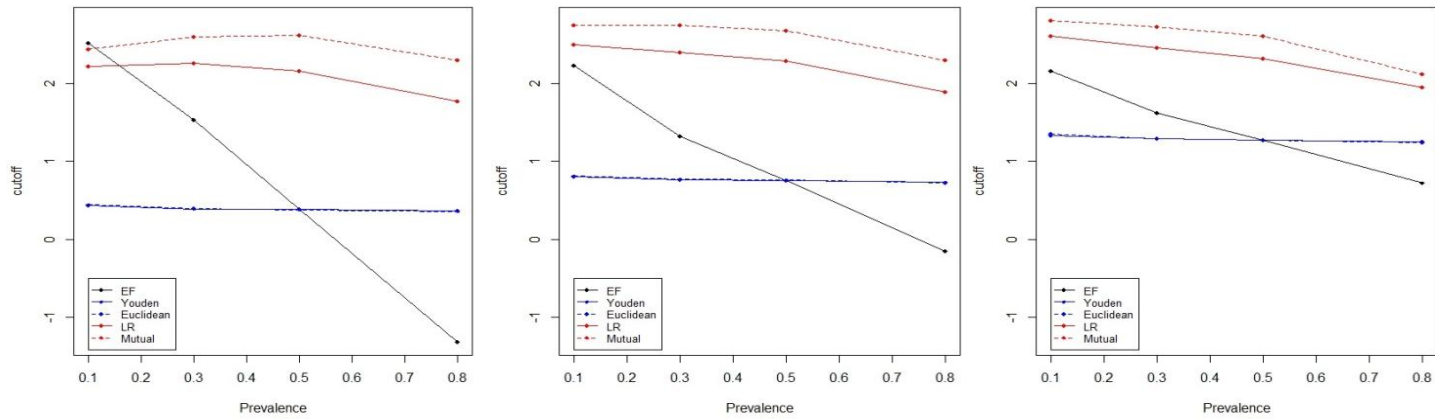


그림 A_2. Random effect (# of 3), Equal variance, Pattern=LL일 때, 최적 기준점(AUC=0.702, 0.855, 0.961)

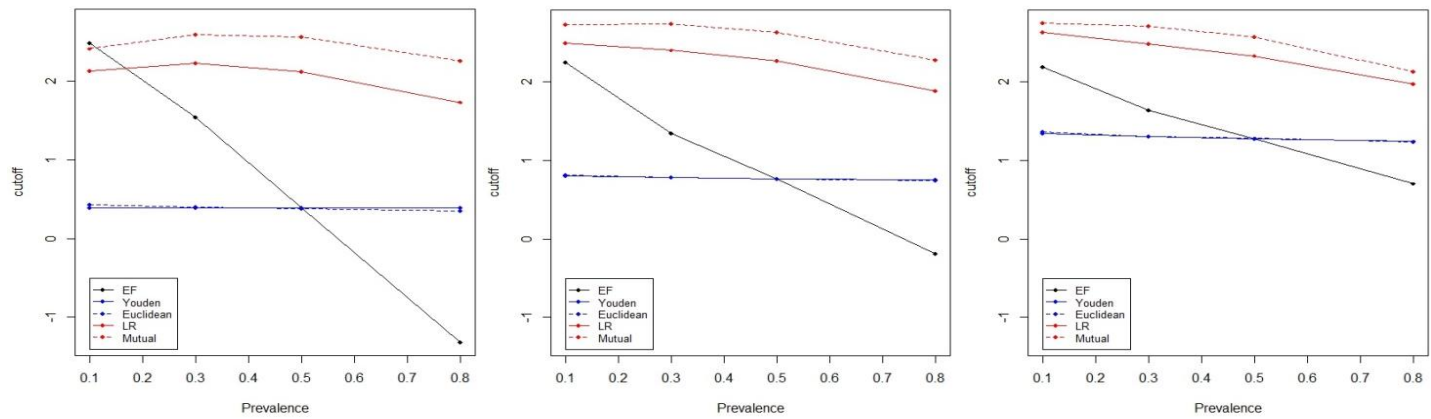


그림 A_3. Random effect (# of 3), Equal variance, Pattern=HH일 때, 최적 기준점(AUC=0.702, 0.855, 0.961)

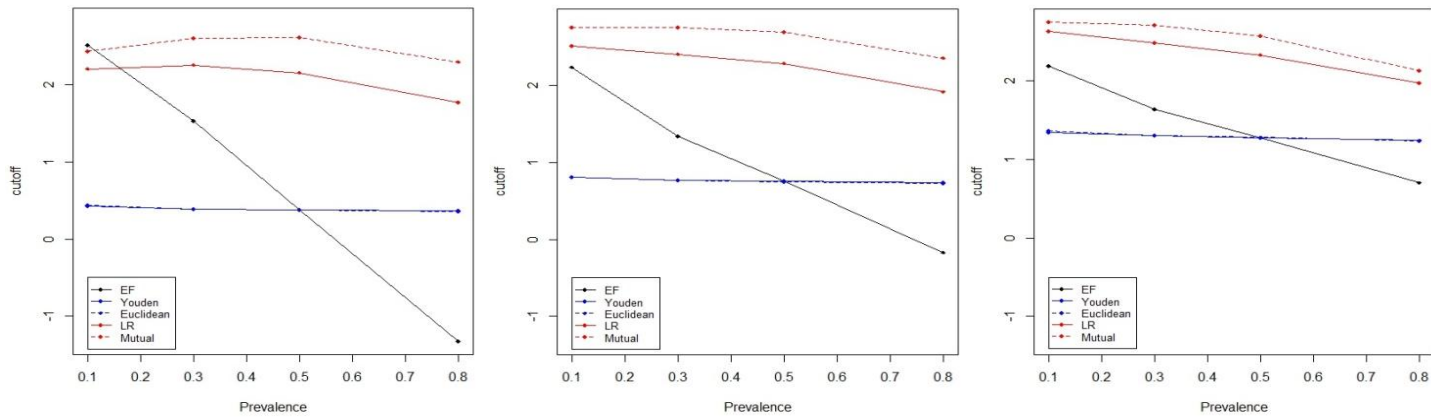


그림 A_4. Random effect (# of 3), Equal variance, Pattern=LH일 때, 최적 기준점 (AUC=0.702, 0.855, 0.961)

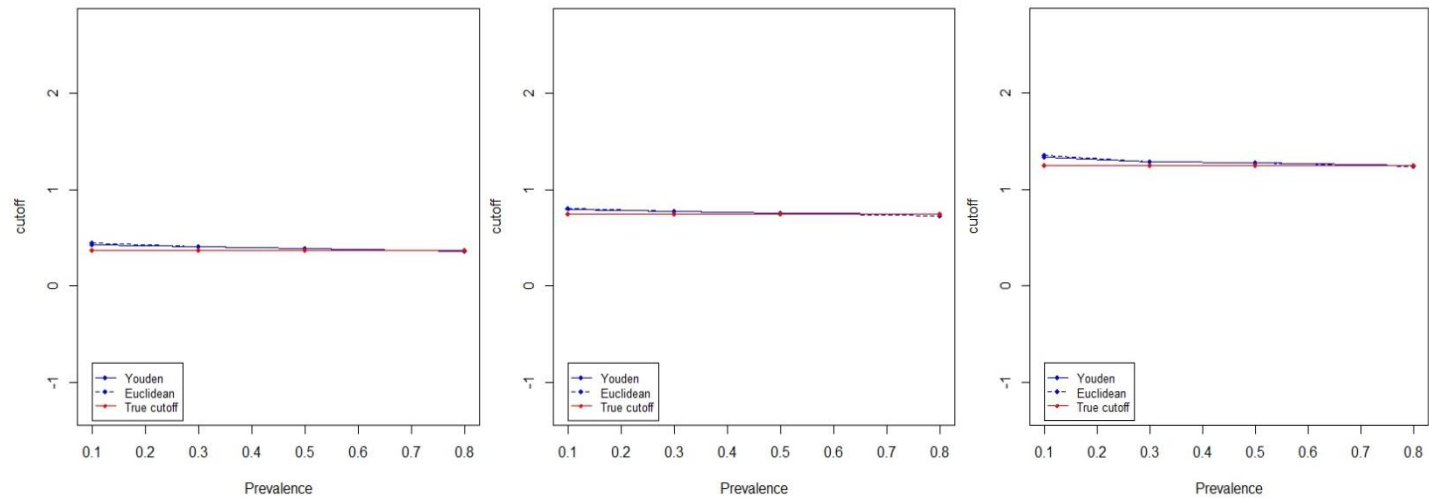


그림 B. Random effect (# of 3), Equal-variance, Pattern=HH일 때,
두 방법(Youden index, Euclidean)의 최적 기준점과 실제 기준점(AUC=0.702, 0.855, 0.961)

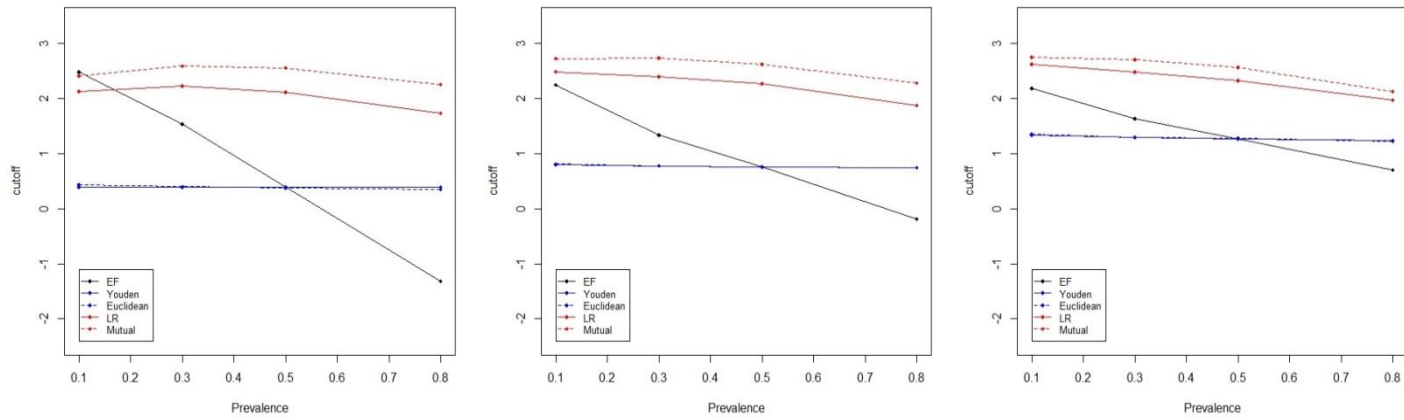


그림 C_1. Random effect (# of 3), Equal variance, Pattern=HH일 때, 최적 기준점 (AUC=0.702, 0.855, 0.961)

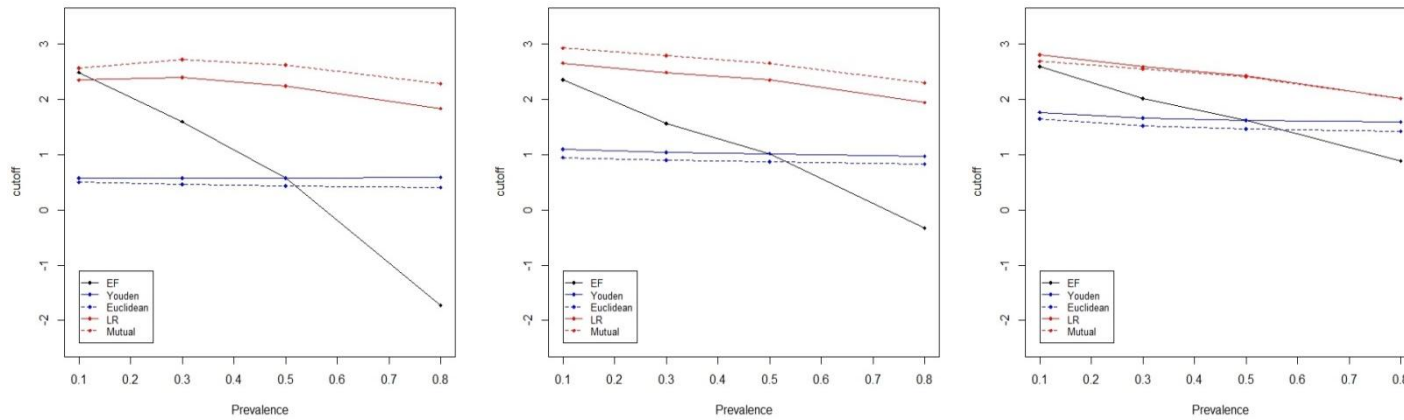


그림 C_2. Random effect (# of 3), Unequal variance, Pattern=HH일 때, 최적 기준점 (AUC=0.702, 0.855, 0.961)

표 A.1. 1000번의 모의실험을 통한 각 방법의 최적의 기준점 비교

- Fixed effect (# of 3), Equal variance, Pattern=HH

AUC	Prevalence	Efficiency	Youden	Euclidean	Likelihood Ratio	Mutual Information
0.702	0.1	3.18 (0.4098)	0.43 (0.5862)	0.47 (0.3096)	2.75 (0.5805)	2.99 (0.3866)
	0.3	2.21 (0.5374)	0.42 (0.4833)	0.42 (0.2199)	2.82 (0.3737)	3.21 (0.3062)
	0.5	0.41 (0.4467)	0.41 (0.4542)	0.39 (0.2097)	2.67 (0.3499)	3.15 (0.3717)
	0.8	-2.07 (0.4920)	0.37 (0.5159)	0.34 (0.2476)	2.25 (0.3715)	2.86 (0.4729)
0.855	0.1	2.94 (0.4359)	0.84 (0.4449)	0.85 (0.2661)	3.05 (0.3367)	3.31 (0.3143)
	0.3	1.82 (0.3401)	0.79 (0.3293)	0.79 (0.1867)	2.97 (0.2785)	3.34 (0.3819)
	0.5	0.78 (0.3126)	0.78 (0.3149)	0.77 (0.1757)	2.81 (0.3095)	3.23 (0.4446)
	0.8	-0.82 (0.3803)	0.75 (0.3616)	0.73 (0.2114)	2.40 (0.5326)	2.89 (0.5089)
0.962	0.1	2.70 (0.3172)	1.35 (0.3615)	1.37 (0.2533)	3.17 (0.2907)	3.44 (0.3837)
	0.3	1.89 (0.2579)	1.28 (0.2784)	1.29 (0.1796)	3.04 (0.2922)	3.38 (0.4142)
	0.5	1.27 (0.2460)	1.27 (0.2487)	1.27 (0.1610)	2.88 (0.3231)	3.26 (0.4681)
	0.8	0.28 (0.2690)	1.26 (0.2889)	1.25 (0.1885)	2.47 (0.3794)	2.90 (0.5303)

표 A_2. 1000번의 모의실험을 통한 각 방법의 최적 기준점에서의 민감도와 특이도

- Fixed effect (# of 3), Equal variance, Pattern=HH

AUC	Efficiency			Youden		Euclidean		Likelihood Ratio		Mutual Information	
	P	Se	Sp	Se	Sp	Se	Sp	Se	Sp	Se	Sp
0.702	0.1	0.0523	0.9958	0.6733	0.6229	0.6457	0.6371	0.1010	0.9775	0.0399	0.9905
	0.3	0.1833	0.9553	0.6361	0.6288	0.6280	0.6273	0.0722	0.9890	0.0266	0.9947
	0.5	0.6232	0.6387	0.6239	0.6380	0.6252	0.6288	0.0832	0.9874	0.0333	0.9932
	0.8	0.9886	0.0858	0.6226	0.6486	0.6307	0.6291	0.1406	0.9729	0.0605	0.9832
0.855	0.1	0.1942	0.9911	0.7609	0.7283	0.7435	0.7362	0.1443	0.9936	0.0786	0.9958
	0.3	0.4461	0.9226	0.7342	0.7273	0.7285	0.7276	0.1485	0.9941	0.0849	0.9952
	0.5	0.7235	0.7341	0.7233	0.7343	0.7253	0.7277	0.1746	0.9923	0.1040	0.9933
	0.8	0.9692	0.2987	0.7200	0.7427	0.7256	0.7305	0.2638	0.9813	0.1618	0.9833
0.962	0.1	0.5050	0.9855	0.8671	0.8422	0.8546	0.8488	0.3376	0.9960	0.2457	0.9963
	0.3	0.7032	0.9317	0.8480	0.8363	0.8411	0.8392	0.3621	0.9951	0.2676	0.9952
	0.5	0.8396	0.8397	0.8396	0.8397	0.8379	0.8384	0.4026	0.9932	0.3011	0.9933
	0.8	0.9643	0.5966	0.8314	0.8492	0.8349	0.8417	0.5155	0.9831	0.3941	0.9833

표 B_1. 1000번의 모의실험을 통한 각 방법의 최적의 기준점 비교

- Fixed effect (# of 3), Unequal variance, Pattern=HH

AUC	Prevalence	Efficiency	Youden	Euclidean	Likelihood Ratio	Mutual Information
0.702	0.1	3.19(0.4415)	0.63(0.6091)	0.54(0.3251)	2.93(0.4943)	3.13(0.3890)
	0.3	2.16(0.4772)	0.66(0.5024)	0.47(0.2325)	2.94(0.3162)	3.29(0.3493)
	0.5	0.65(0.4516)	0.65(0.4601)	0.46(0.2101)	2.77(0.3289)	3.20(0.4150)
	0.8	-2.42(0.5893)	0.59(0.5241)	0.42(0.2587)	2.36(0.3592)	2.88(0.4964)
0.855	0.1	2.91(0.3943)	1.17(0.4629)	1.03(0.3158)	3.18(0.3174)	3.42(0.3761)
	0.3	2.01(0.3078)	1.11(0.3370)	0.95(0.2085)	3.05(0.2960)	3.38(0.4172)
	0.5	1.12(0.3223)	1.12(0.3245)	0.94(0.1993)	2.88(0.3315)	3.25(0.4706)
	0.8	-0.91(0.5887)	1.06(0.3724)	0.88(0.2309)	2.46(0.3897)	2.90(0.5150)
0.962	0.1	2.99(0.3339)	1.89(0.4239)	1.77(0.3604)	3.26(0.3122)	3.34(0.5860)
	0.3	2.30(0.2640)	1.82(0.3048)	1.65(0.2228)	3.10(0.3094)	3.32(0.5046)
	0.5	1.78(0.2769)	1.77(0.2799)	1.61(0.2047)	2.92(0.3363)	3.16(0.5592)
	0.8	0.70(0.3683)	1.75(0.3006)	1.58(0.2264)	2.50(0.3955)	2.66(0.6127)

표 B.2. 1000번의 모의실험을 통한 각 방법의 최적 기준점에서의 민감도와 특이도

- Fixed effect (# of 3), Unequal variance, Pattern=HH

AUC	Efficiency			Youden		Euclidean		Likelihood Ratio		Mutual Information	
	P	Se	Sp	Se	Sp	Se	Sp	Se	Sp	Se	Sp
0.702	0.1	0.0790	0.9957	0.6355	0.6734	0.6394	0.6543	0.0971	0.9874	0.0501	0.9933
	0.3	0.2235	0.9535	0.5848	0.6925	0.6219	0.6435	0.0847	0.9928	0.0436	0.9950
	0.5	0.5747	0.6995	0.5752	0.6990	0.6164	0.6481	0.1011	0.9910	0.0540	0.9933
	0.8	0.9897	0.0615	0.5765	0.7063	0.6178	0.6526	0.1580	0.9791	0.0874	0.9833
0.855	0.1	0.3184	0.9904	0.7295	0.8015	0.7462	0.7754	0.2309	0.9957	0.1656	0.9962
	0.3	0.4927	0.9434	0.7017	0.8025	0.7307	0.7665	0.2484	0.9951	0.1844	0.9952
	0.5	0.6899	0.8105	0.6901	0.8103	0.7248	0.7688	0.2778	0.9932	0.2066	0.9933
	0.8	0.9575	0.2943	0.6906	0.8138	0.7287	0.7666	0.3606	0.9831	0.2693	0.9833
0.962	0.1	0.6924	0.9924	0.8728	0.9194	0.8807	0.9057	0.6189	0.9963	0.5931	0.9934
	0.3	0.7768	0.9671	0.8490	0.9186	0.8653	0.8974	0.6251	0.9952	0.5751	0.9949
	0.5	0.8436	0.9202	0.8441	0.9197	0.8627	0.8962	0.6522	0.9933	0.6027	0.9931
	0.8	0.9401	0.7133	0.8388	0.9265	0.8602	0.8985	0.7213	0.9833	0.6910	0.9831

표 C.1. 1000번의 모의실험을 통한 각 방법의 최적의 기준점 비교

- Random effect (# of 10), Equal variance, Pattern=HH

AUC	Prevalence	Efficiency	Youden	Euclidean	Likelihood Ratio	Mutual Information
0.702	0.1	2.48(0.3704)	0.42(0.4181)	0.44(0.2375)	2.11(0.5536)	2.40(0.3268)
	0.3	1.59(0.5045)	0.40(0.3420)	0.40(0.1823)	2.23(0.3927)	2.60(0.2872)
	0.5	0.39(0.3157)	0.39(0.3195)	0.39(0.1662)	2.13(0.3495)	2.56(0.3440)
	0.8	-1.30(0.4812)	0.37(0.3608)	0.35(0.1923)	1.76(0.3689)	2.25(0.4575)
0.855	0.1	2.26(0.3906)	0.81(0.3329)	0.82(0.2332)	2.47(0.3331)	2.71(0.3089)
	0.3	1.37(0.2664)	0.77(0.2474)	0.78(0.1646)	2.39(0.3125)	2.72(0.4029)
	0.5	0.76(0.2315)	0.76(0.2331)	0.76(0.1619)	2.27(0.3007)	2.64(0.4294)
	0.8	-0.18(0.3109)	0.73(0.2707)	0.73(0.1830)	1.87(0.3536)	2.27(0.4982)
0.962	0.1	2.20(0.3019)	1.34(0.3044)	1.36(0.2595)	2.61(0.3197)	2.73(0.5630)
	0.3	1.65(0.2202)	1.28(0.2345)	1.29(0.1823)	2.47(0.3348)	2.78(0.5576)
	0.5	1.27(0.2005)	1.27(0.2018)	1.27(0.1670)	2.34(0.3475)	2.57(0.5591)
	0.8	0.71(0.2541)	1.24(0.2381)	1.23(0.1951)	1.95(0.3973)	2.08(0.6451)

표 C.2. 1000번의 모의실험을 통한 각 방법의 최적 기준점에서의 민감도와 특이도

- Random effect (# of 10), Equal variance, Pattern=HH

AUC	Efficiency			Youden		Euclidean		Likelihood Ratio		Mutual Information	
	P	Se	Sp	Se	Sp	Se	Sp	Se	Sp	Se	Sp
0.702	0.1	0.0784	0.9943	0.7126	0.6571	0.6876	0.6689	0.1444	0.9664	0.0506	0.9890
	0.3	0.2682	0.9370	0.6708	0.6608	0.6616	0.6617	0.1000	0.9834	0.0366	0.9940
	0.5	0.6617	0.6680	0.6621	0.6677	0.6600	0.6624	0.1080	0.9848	0.0451	0.9930
	0.8	0.9803	0.1478	0.6541	0.6933	0.6630	0.6738	0.1805	0.9737	0.0892	0.9832
0.855	0.1	0.3052	0.9877	0.8178	0.7794	0.8006	0.7881	0.2146	0.9926	0.1216	0.9954
	0.3	0.5848	0.9170	0.7913	0.7801	0.7848	0.7819	0.2192	0.9935	0.1377	0.9951
	0.5	0.7854	0.7836	0.7852	0.7838	0.7835	0.7817	0.2442	0.9921	0.1538	0.9933
	0.8	0.9589	0.4620	0.7854	0.7972	0.7841	0.7926	0.3710	0.9814	0.2515	0.9833
0.962	0.1	0.6723	0.9864	0.9212	0.8981	0.9112	0.9037	0.5003	0.9960	0.4394	0.9925
	0.3	0.8184	0.9507	0.9034	0.8955	0.8986	0.8974	0.5276	0.9951	0.4458	0.9933
	0.5	0.8983	0.8998	0.8982	0.8999	0.8974	0.8985	0.5668	0.9932	0.4803	0.9918
	0.8	0.9681	0.7702	0.8980	0.9099	0.9001	0.9048	0.6983	0.9833	0.6410	0.9807

표 D_1. 1000번의 모의실험을 통한 각 방법의 최적의 기준점 비교

- Random effect (# of 10), Unequal variance, Pattern=HH

AUC	Prevalence	Efficiency	Youden	Euclidean	Likelihood Ratio	Mutual Information
0.702	0.1	2.48(0.3975)	0.60(0.4575)	0.51(0.2608)	2.35(0.4523)	2.58(0.3407)
	0.3	1.59(0.3884)	0.61(0.3638)	0.46(0.1937)	2.38(0.3211)	2.70(0.3514)
	0.5	0.60(0.3317)	0.60(0.3345)	0.44(0.1737)	2.25(0.3388)	2.62(0.4033)
	0.8	-1.74(0.6015)	0.56(0.3842)	0.41(0.2165)	1.86(0.3745)	2.27(0.4975)
0.855	0.1	2.35(0.3726)	1.08(0.3777)	0.95(0.2818)	2.66(0.3458)	2.90(0.4252)
	0.3	1.59(0.210)	1.03(0.2833)	0.89(0.2011)	2.49(0.3357)	2.77(0.4459)
	0.5	1.02(0.2603)	1.02(0.2610)	0.87(0.1810)	2.35(0.3613)	2.66(0.4552)
	0.8	-0.34(0.5592)	0.99(0.3106)	0.83(0.2260)	1.95(0.4148)	2.28(0.5250)
0.962	0.1	2.61(0.3789)	1.75(0.4209)	1.64(0.3969)	2.79(0.3706)	2.63(0.5941)
	0.3	2.01(0.3062)	1.65(0.3192)	1.51(0.2710)	2.58(0.3858)	2.52(0.5922)
	0.5	1.64(0.2955)	1.64(0.2986)	1.48(0.2487)	2.44(0.3986)	2.41(0.5459)
	0.8	0.87(0.3514)	1.57(0.3369)	1.42(0.2836)	2.03(0.4492)	2.02(0.5434)

표 D_2. 1000번의 모의실험을 통한 각 방법의 최적 기준점에서의 민감도와 특이도

- Random effect (# of 10), Unequal variance, Pattern=HH

AUC	Efficiency			Youden		Euclidean		Likelihood Ratio		Mutual Information	
	P	Se	Sp	Se	Sp	Se	Sp	Se	Sp	Se	Sp
0.702	0.1	0.1276	0.9934	0.6628	0.7140	0.6707	0.6929	0.1356	0.9839	0.0670	0.9926
	0.3	0.3061	0.9429	0.6093	0.7308	0.6467	0.6830	0.1167	0.9926	0.0669	0.9949
	0.5	0.6011	0.7346	0.6007	0.7350	0.6442	0.6811	0.1342	0.9908	0.0773	0.9932
	0.8	0.9847	0.0930	0.5992	0.7522	0.6474	0.6919	0.2088	0.9797	0.1325	0.9833
0.855	0.1	0.4276	0.9897	0.7696	0.8453	0.7864	0.8198	0.3250	0.9958	0.2499	0.9960
	0.3	0.5941	0.9457	0.7410	0.8475	0.7690	0.8124	0.3417	0.9951	0.2706	0.9952
	0.5	0.7336	0.8511	0.7337	0.8510	0.7653	0.8119	0.3720	0.9932	0.2949	0.9933
	0.8	0.9410	0.4322	0.7323	0.8646	0.7687	0.8185	0.4734	0.9832	0.3882	0.9833
0.962	0.1	0.7727	0.9940	0.9023	0.9435	0.9093	0.9323	0.7204	0.9963	0.7455	0.9903
	0.3	0.8335	0.9762	0.8833	0.9434	0.8967	0.9260	0.7288	0.9952	0.7359	0.9921
	0.5	0.8786	0.9452	0.8788	0.9450	0.8941	0.9252	0.7508	0.9933	0.7502	0.9913
	0.8	0.9432	0.8134	0.8778	0.9523	0.8947	0.9292	0.8106	0.9833	0.8103	0.9819

표 E. 1000번의 모의실험을 통한 두 방법 간 유병률에 따른 최적 기준점과 실제 기준점의 차이(bias)

- Random effect (# of 3), Equal variance, AUC=0.961)

Pattern	P=0.1		P=0.3		P=0.5		P=0.8	
	Youden	Euclidean	Youden	Euclidean	Youden	Euclidean	Youden	Euclidean
HL	7.2	8.8	3.2	3.2	2.4	2.4	0	-0.8
LL	6.4	8	3.2	3.2	1.6	1.6	0	-0.8
HH	7.2	8.8	4	4	1.6	2.4	-0.8	-1.6
LH	5.6	7.2	4	4	1.6	1.6	0	-0.8

ABSTRACT

Comparison of methods for finding the optimal cutoff value in MRMC data

In the medical field, various diagnostic tools such as medical images or blood tests are used to diagnostic test. Is it also essential to select an optimal cutoff that maximizes disease predictability for useful in a clinical setting.

Particularly, there is a possibility that the reader's subjectivity will be involved in the case of using the image reading result for diagnosis of disease. Therefore, this paper uses the results of combined disease diagnostic test result considering various deviations. As a result of the combined disease diagnostic test result, the accuracy evaluation of the diagnostic tool and the selection of the optimal cutoff are performed using the Multi-Reader Multi-Case (MRMC) ROC Curve. MRMC ROC Curve analysis methods are parametric (Gallas, 2006) and nonparametric based on ANOVA. The DBM model (Dorman, Berbaum, and Metz, 1992) and OR model (Obuchowski and Rockette, 2000) are used for the nonparametric method.

MRMC ROC Curve study has been actively performed. However, there are no previous proposals and comparative studies on how to select the optimal cutoff in MRMC ROC Curve. Accordingly, this study compared the methods (Efficiency, Youden index, Euclidean, Likelihood Ratio, and Mutual Information). The methods of finding the optimal cutoff in the ROC Curve was applied to the MRMC ROC Curve.

The simulations of this study are based on various situations and examined the optimal cutoff according to the method. As a result, it was confirmed that the optimal cutoff value was selected in a similar pattern in the two assumptions (equal variance, unequal variance in between disease group and nondisease group) of two models (fixed effect, random effect about reader). In Youden index and Euclidean method, the stable optimal cutoff was selected regardless of prevalence.

Therefore, if there are significant differences between the prevalence of the population and the prevalence of the sample population, we recommend using the method Youden index and Euclidean. In other cases, it is recommended to select a method according to the researcher's research purpose.

Key words : Optimal cutoff, MRMC ROC Curve,
Efficiency, Youden index, Euclidean,
Likelihood Ratio, Mutual Information