



# R을 이용한 결과 변수에 따른 최적의 Cut-off Point 결정

양주연, 이해선

연세대학교 의과대학 의학통계지원실

## Determining the Optimal Cut-off Point According to the Outcome Variables Using R

Juyeon Yang, Hye Sun Lee

Biostatistics Collaboration Unit, Yonsei University College of Medicine, Seoul, Korea

Received October 28, 2022  
Accepted November 12, 2022

Corresponding author:

Hye Sun Lee

E-mail: [hslee1@yuhs.ac](mailto:hslee1@yuhs.ac)

<https://orcid.org/0000-0001-6328-6948>

Clinical research ultimately aimed to promptly diagnose and prevent diseases through precise biomarker development. Finding the optimal cut-off point of a regularly measured biomarker can help its interpretation and ultimately help in disease investigation and diagnosis, more specifically in determining the presence of diseases. Therefore, this study aimed to use the characteristics of outcome variables in clinical research to explain how to determine the optimal cut-off point. The outcome variables can be divided into dichotomous, ordinal, and survival types. The optimal cut-off point can be determined by finding points that maximize the Youden index, extended Youden index, and log-rank statistics. This study will enable clinical researchers to accurately determine the optimal cut-off points for regularly measured biomarkers, thereby enabling prompt disease diagnosis for effective treatment.

**Key Words:** Optimal cut-off point; Youden index; Log-rank statistics; Biomarkers

### INTRODUCTION

임상 연구에서 항상 강조되고 목표로 생각하는 바는 만성 또는 중증 질병을 조기 진단하고, 예방하는 것이다. 이러한 목표는 새롭고 정밀한 바이오마커의 개발로 이어진다[1]. 실제로 임상에서 질병의 유무는 바이오마커 수치 평가를 통해 판별될 수 있으며, 바이오마커를 해석하는 것은 질병을 조사하고 진단하는 면에서 중요한 역할을 하고 있다. 바이오마커를 통한 질병 진단은 바이오마커의 수치와 질병 상태 간에 상관성을 가지고 있는 것을 기반으로 한다. 예를 들어, 특정 질병을 가진 집단의 바이오마커 수치는 질병이 없는 집단에 비해 보통 높을 수 있다는 것이다. 질병 유무를 분류하는 데 바이오마커를 활용하기 위해, 연속형 바이오마커의 최적 cut-off point를 설정

하고, cut-off point보다 큰 쪽은 질병을 갖는 쪽으로, 작은 쪽은 질병을 가지지 않는 쪽으로 분류할 수 있다[2]. 바이오마커의 cut-off point에 따라 질병 유무가 달라질 수 있으므로 정확하게 판단할 수 있는 cut-off point를 찾아야 하며, 검증을 통해 정확도를 확인을 해야 한다.

본 논문에서는 결과 변수의 유형에 따른 최적의 cut-off point를 끊는 방법을 설명함으로써 임상연구자가 질병 진단을 효율적이고 정확하게 수행할 수 있는 기반을 만드는 것을 목표로 한다. 다음 장에서는 이분형 결과 변수, 순위형 결과 변수, 생존형 결과 변수에서 각각 최적의 cut-off point를 끊는 원리와 방법을 알아보려고 한다. 또한, cut-off point를 끊을 때 주의해야 할 사항에 대하여 소개하고자 한다. 또한 모든 과정은 R을 통해 구현해 낼 수 있도록 R 코드를 제공하며, 데이터는 “survival”



패키지의 “colon” 데이터, “DiagTest3Grp” 패키지의 “AL”, “maxstat” 패키지의 “DLBCL” 데이터를 사용하였다.

## MAIN SUBJECTS

### 이분형 결과 변수에 대한 Cut-off Point

#### Youden index (J)의 정의 및 개념

질병 유무와 같은 이분형 결과 변수를 예측하기 위한 연속형 인자의 cut-off point를 찾을 때 가장 많이 사용하는 방법은 Youden index (J)를 이용하는 것이다[3]. 이는 1950년에 Youden에 의해서 처음 제안된 방법으로, 질병 여부와 진단 도구의 결과를 이용한 분류표를 통해 의미를 부여하고 식을 만들었다. 질병을 진단하는데 있어 진단 도구를 통해 질병이 있는 사람은 진단 도구 결과가 양성 나오기를 바라고 질병이 없는 사람은 진단 도구 결과가 음성이 나오기를 바란다. 이 원리를 이용하여 질병이 있는 집단 내 양성이라고 제대로 판단한 사람에서 음성이라고 잘못 판단한 사람을 빼고, 질병이 없는 집단 내 음성이라고 제대로 판단한 사람에서 양성이라고 잘못 판단한 사람을 빼어 동일한 가중치를 주고 합을 구한 것이

**Table 1.** Contingency table of dichotomous disease outcome and test result

Disease	Test		Total
	Positive (+)	Negative (-)	
Positive (+)	a	b	a + b
Negative (-)	c	d	c + d
Total	a + c	b + d	a + b + c + d

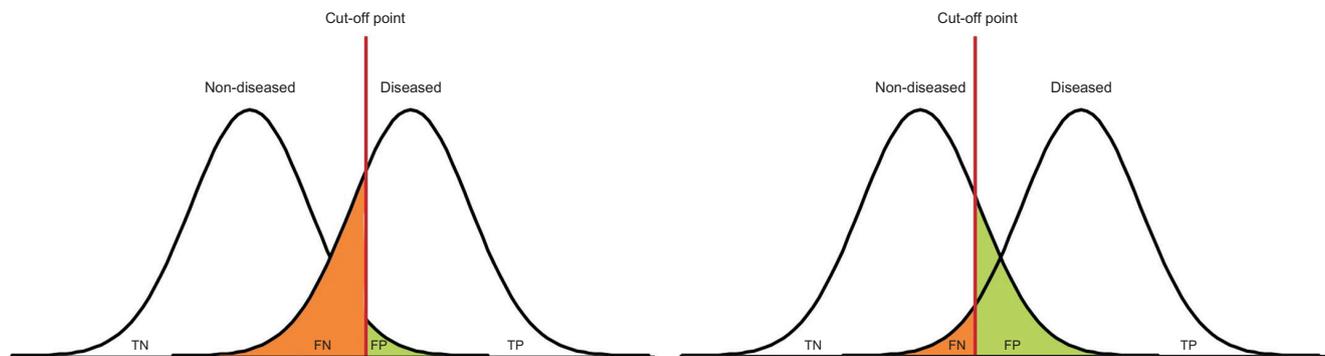
Youden index (J)이다. 이는 흔히 말하는 민감도와 특이도의 합에 1을 빼준 것과 같은 결과가 된다.

#### 방법 및 특징

Youden index는 Table 1에 따라 식 (1)을 이용하여 구할 수 있다. Youden index는 0과 1 사이에 존재하는 값으로 계산이 되며, 이 값이 0이라는 것은 질병이 없는 집단과 질병이 있는 집단에서 양성 비율이 같음을 의미하고, 이 경우 진단 도구의 가치가 없다고 볼 수 있다. Youden index가 1이라는 것은 질병이 없는 집단에서 양성 비율과 질병이 있는 집단에서 음성비율이 모두 높다는 것으로, 진단 도구로써 유용한 가치가 있다고 볼 수 있다. Youden index는 상대적인 크기 비교도 가능하며, 기존 검사 도구에 비해 새로운 검사 도구의 Youden index가 크다면 민감도와 특이도를 통합한 측면에서 판별능력이 높다고 볼 수 있다.

$$\text{Youden index (J)} = \frac{1}{2} \left( \frac{a-b}{a+b} + \frac{d-c}{c+d} \right) = \frac{a}{a+b} + \frac{d}{c+d} - 1 = \frac{ad-bc}{(a+b)(c+d)} \quad (1)$$

Fig. 1과 같이 만약 cut-off point를 질병이 없는 집단에 중점을 두게 되면 특이도가 커지는 반면, 질병이 없음에도 질병이 있다고 판단하는 위음성이 증가하게 되며, 질병이 있는 집단에 중점을 두게 되면 민감도가 커지는 반면, 질병이 없음에도 불구하고 질병이 있다고 판단하는 위양성이 증가하게 된다. 위음성이나 위양성이 너무 커지는 것을 방지하기 위하여, 연속형 자료의 모든 값을 cut-off point로 설정하여 이분형으로 나누어 가며 Youden index (J)를 계산하여 최대가 되는 점을 최적의 cut-off point 값으로 결정하게 된다. 한편, receiver operating characteristic (ROC) 곡선의 면적인 area under the

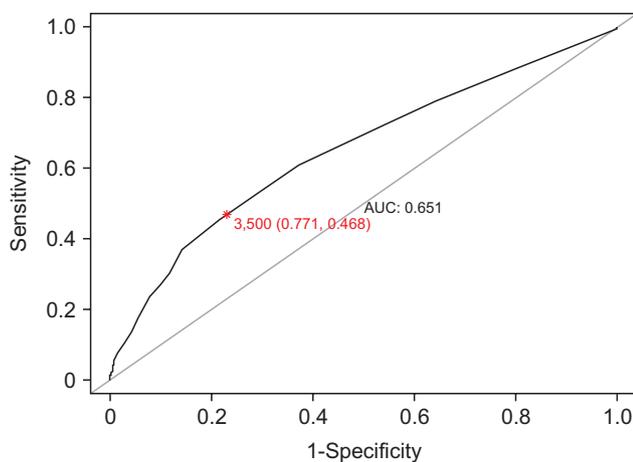


**Fig. 1.** False positive and false negative according to cut-off point. TN, true negative; FN, false negative; FP, false positive; TP, true positive.

curve (AUC)는 (민감도+특이도)/2로 Youden index (J)와 비례하는 값이므로, ROC 곡선을 활용하여 AUC가 최대가 되는 점을 찾는 방법으로 구한 cut-off point 값과 Youden index (J)가 최대가 되는 점으로 구한 cut-off point 값은 같게 된다.

### 예시

R에서는 “pROC” 패키지를 이용하여 Youden 방법을 이용한 최적의 cut-off point를 구할 수 있다[4]. “Colon” 자료에서 대장암 환자의 사망을 진단할 수 있는 림프절 개수의 cut-off point를 찾고자 할 때, 먼저 “roc” 함수를 이용하여 ROC 곡선을 생성한다. Response에 결과 변수인 “status”를 넣고, predictor에 독립 변수인 “nodes”를 넣는다. Levels 조건은 결과 변수의 형태를 지정하는 것으로, “status” 변수는 0과 1로 되어 있으므로 “c(0,1)”로 지정하고, 사망인 1을 예측하고자 하므로 direction은 “<”으로 설정한다. “Coords” 함수는 구축된 ROC 곡선에서 Youden 방법을 사용하여 최적의 cut-off point를 결정한다. 이를 통해 3.5가 적정 cut-off point이며, 이 때 AUC는 0.651임을 알 수 있으며, Fig. 2를 통해 ROC 곡선과 cut-off point를 확인할 수 있다. Jang et al. [5]를 통해 이분형 결과 변수에서 Youden index를 통해 cut-off point를 찾는 사례를 참고할 수 있다.



**Fig. 2.** ROC curve and optimal cut-off point. ROC, receiver operating characteristic; AUC, area under the curve.

[R code]

```
>library(pROC)
>library(survival)
>data<-colon
>AUC<-roc(response=data$status,
predictor=data$nodes, levels=c(0,1),direction=
"<")
>coords(AUC, x="best", best.method="youden")
>plot.roc(AUC, print.auc=TRUE, print.
thres=TRUE, print.thres.pch=8,print.thres.col=
"red", legacy.axes =TRUE)
```

### 기타 방법

이분형 결과 변수에 대하여 cut-off point를 끊는 방법은 Youden index 외에도 다양한 방법들이 있다[6]. Euclidean 방법은 (0, 1)과 가장 가까운 ROC 곡선 위의 점을 cut-off point로 설정하는 방법이다. Likelihood ratios을 이용한 방법은 두 가지로 나눌 수 있는데 Likelihood positive (LR+)는 위양성 증 진양성의 비율이 최대가 되도록 cut-off point로 끊는 방법이며, Likelihood negative (LR-)는 진음성 중 위음성의 비율이 최소가 될 수 있는 cut-off point를 찾는 것이다[7]. Efficiency 방법은 질병이 있는 대상자들이 진단결과 양성으로 판단되는 경우와 질병이 없는 대상자들이 진단결과 음성으로 판단되는 경우의 합을 전체 대상자 수로 나누어 주는 값을 최대로 만드는 cut-off point를 찾는 방법이다. 이 방법은 질병의 유병률에 영향을 많이 받으며, 질병의 유병률이 감소할수록 검사의 정확성이 증가하는 경향이 있다. Misclassification costs term 방법은 위음성과 위양성으로 잘못 분류될 비용을 알고 있을 때, 유병률과 민감도, 특이도를 이용하여 계산한 값이 최소가 될 수 있는 cut-off point를 찾는 방법이다. 앞서 설명한 기타 방법을 이용하여 최적의 cut-off point를 끊는 원리는 Table 2와 같다. Youden index와 Euclidean 방법은 유병률과 관계없이 일정한 최적의 cut-off point를 선정하기 때문에 모집단의 유병률과 표본집단의 유병률의 편차가 심한 경우에 적절할 수 있다. 이에 따라 상황에 맞게 연구자가 적절히 방법을 선택하여 이분형 결과 변수에 대한 cut-off point를 끊을 수 있다.

**Table 2.** Various method of finding optimal cut-off point for dichotomous outcome

Method	Principle
Euclidean method	Minimize $\sqrt{(1 - sensitivity)^2 + (1 - specificity)^2}$
Likelihood ratios	LR+: Maximize $\frac{sensitivity}{1 - specificity}$ LR-: Minimize $\frac{1 - sensitivity}{specificity}$
Efficiency	Maximize $p * sensitivity + (1 - p) * specificity$ , $p = prevalence\ rate$
Misclassification costs term	Minimize $r * p * (1 - sensitivity) + (1 - p) * (1 - specificity)$ , $r = \frac{Cost(FN)}{Cost(FP)}$ , $p = prevalence\ rate$

**Table 3.** Contingency table of ordinal disease outcome and test result

Disease	Test			Total
	Severe	Moderate	Mild	
Severe	a	b	c	a + b + c
Moderate	d	e	f	d + e + f
Mild	g	h	i	g + h + i
Total	a + d + g	b + e + h	c + f + i	a + b + ... + h + i

**순위형 결과 변수에 대한 Cut-off Point**

**확장된 Youden index의 정의 및 개념**

결과 변수가 질병 여부와 같은 이분형 결과 변수가 아닌 경증, 중등증, 중증과 같이 순위형인 경우에는 앞서 다루었던 Youden index (J)를 확장하여 적용할 수 있다 [8-10]. 확장된 Youden index (J)는 진단 도구 결과를 경증인 군을 경증으로, 중등증인 군을 중등증으로, 중증인 군을 중증으로 올바르게 판단한 비율의 합을 구함으로써 결과 변수가 이분형인 자료의 확장 형태이다. 확장된 Youden index를 이용하여 경증 군과 중등증 군을 비교하는 최적의 cut-off point와 중등증 군과 중증 군을 비교하는 최적의 cut-off point를 구할 수 있다. 확장된 Youden index (J)는 Table 3를 이용하여 식 (2)와 같이 구할 수 있으며, se는 민감도와 비슷한 개념으로 중증인 그룹을 중증으로 잘 분류할 확률을, sm은 중등증인 그룹을 중등증으로 잘 분류할 확률을, sp는 특이도와 비슷한 개념으로 경증인 그룹을 경증으로 잘 분류할 확률을 의미한다. 이 값 역시 0과 1사이의 값을 가지며, 1에 가까울수록 판별 능력이 높다고 할 수 있고, 0에 가까울수록 판별 능력이 낮다고 해석할 수 있다. 최적의 cut-off point는 확장된 Youden index (J)가 가장 커지는 점으로 결정한다.

Extended Youden Index (J) =

$$\frac{1}{2} \left( \frac{a}{a+b+c} + \frac{e}{d+e+f} + \frac{i}{g+h+i} - 1 \right) = \frac{1}{2} (se + sm + sp - 1) \quad (2)$$

**예시**

순위형 결과 변수에 대한 cut-off point를 찾는 것은 “DiagTest3Grp” 패키지에서 수행할 수 있다. 패키지 안의 “AL” 데이터를 이용하여 알츠하이머를 D+, D0, D-로 진단함에 있어 신경심리학적 테스트의 측정치의 cut-off point를 정하려고 한다. 먼저 알츠하이머 그룹에 따른 테스트 측정치를 따로 저장하고, “Youden3Grp” 함수를 이용하여 세 개의 집단에 따른 값을 넣어준 후, method는 “Normal”로 설정한다. 그 결과 cut-off point는 0.584과 2.904로, 신경심리학적 테스트 측정치가 0.584보다 아래일 경우 경증, 0.584와 2.904 사이일 경우 중등증, 2.904보다 클 경우 중증으로 분류할 수 있으며, 이 때 sp는 0.903, sm은 0.498, se는 0.777임을 확인할 수 있다. Sinh 등[5]을 통해 순위형 결과 변수에서 확장된 Youden index를 통해 cut-off point를 찾는 실제 사례를 참고할 수 있다.

```
[R code]
>library(DiagTest3Grp)
>data(AL)
>group <- AL$group
```

```
>factor1 <- -AL$FACTOR1
>x <- factor1[group=="D-"]; y <- factor1[group=="D0"]; z <- factor1[group=="D+"]
>youden.res <- Youden3Grp(x=x, y=y, z=z, method = "Normal")
>print(youden.res)
```

**생존형 결과 변수에 대한 Cut-off Point**

**로그 순위 통계량의 정의 및 개념**

로그 순위 통계량은 집단의 생존 곡선이 같은지 검정할 때 사용하는 통계량으로 생존형 결과 변수에서 생존 확률을 가장 높게 하는 cut-off point를 찾고자 할 때 이 용할 수 있다[11-13]. 기본적인 원리는 특정 값을 기준으로 그룹을 나눈 후, 두 그룹에서 계산된 로그 순위 통계량이 가장 커지는 특정 값을 최적의 cut-off point로 지정하는 것이다. 로그순위 통계량은 전체 집단을 관측 순으로 배열하고, 절단된 경우를 제외한 후 사건이 발생한 구간들에 대해 군별로 각 구간의 사망에 대한 기대 빈도를

통해 구할 수 있다. 예를 들어 Table 4과 같이 1과 2 집단의 생존 자료가 있다고 가정하자. 첫 번째로, 두 집단을 각각 시간순으로 나열한 후 관측자 수와 생존자 수를 정리한다. 1번 집단에서는 시점 6에서 사건이 발생하였으므로 생존자 수는 4명이며, 관측자 수는 5이다. 이렇게 집단별로 정리가 되었으면, 중도 절단된 경우를 제외하고 두 집단을 합하여 시간순으로 나열한 후, Table 5와 같이 두 그룹의 관측자 수와 사건 관측/기대 빈도를 정리한다. 여기서 기대 빈도는 식 (3)을 통해 계산 가능하다. Table 5에서 시점이 12일 때, 1 집단에서의 관측자 수는 시점 6에서 이미 한 명의 사건이 발생했으므로 4이며, 2 집단에서의 관측자 수는 이전 시점에서 사건이 발생하지 않았으므로 5이다. 이 때 집단 1의 사건 관측 빈도는 1이며, 기대 빈도는  $1 \cdot (4 / (4 + 9))$ 로 0.444이다. 집단 2의 경우, 관측 빈도는 0이며, 기대 빈도는  $5 / (4 + 5)$ 로 0.556이다. 이렇게 모든 시점에서 계산을 하고, 집단 별로 관측 빈도와 기대 빈도의 합을 구한다. 식 (4)를 통해 로그 순위 통계량을 구할 수 있으며, 예제의 경우 로그 순위 통계량은  $\frac{(4-2.997)^2}{2.997} + \frac{(4-4.003)^2}{4.003} = 0.336$  이 된다. 로그 순위 통계량은 그

**Table 4.** Example of survival data

Trt	Time	Event	Number of survivors	Number of observers
1	6	1	4	5
1	9+	0	4	4
1	12	1	3	4
1	24	1	2	3
1	30+	0	2	2
2	18	1	4	5
2	30	1	3	4
2	39	1	2	3
2	43	1	1	2
2	45+	0	1	1

**Table 5.** Observed frequency and expected frequency of two groups

Time	Number of observers		Number of trt1's deaths		Number of trt2's deaths	
	Trt 1	Trt 2	Observed frequency	Expected frequency	Observed frequency	Expected frequency
6	5	5	1	0.5	-	0.5
12	4	5	1	0.444	-	0.556
18	4	5	-	0.444	1	0.556
30	3	5	1	0.375	-	0.625
24	2	4	1	0.333	1	0.667
39	2	3	-	0.4	1	0.6
43	2	2	-	0.5	1	0.5
Total			4	2.997	4	4.003

값이 클수록 두 집단 간 생존 곡선이 다르다고 본다. 그렇기 때문에, 연속형 변수를 특정 값을 기준으로 두 그룹으로 나눈 후 위와 같이 로그 순위 통계량을 계산하여 그 값이 가장 큰 점을 최적의 cut-off point로 결정한다.

$$\text{기대빈도} = \text{관측빈도} \times \frac{\text{각 군의 관측 대상 수}}{\text{각 군의 관측 대상 수의 합}} \quad (3)$$

$$\text{Log rank statistics} = \sum \frac{(\text{관측빈도 합} - \text{기대빈도 합})^2}{\text{기대빈도}} \quad (4)$$

**예시**

생존형 결과 변수에 대한 cut-off point는 “maxstat” 패키지를 이용하여 구할 수 있다[14]. “Maxstat” 패키지에 있는 “DLBCL” 데이터를 이용하여 DLBCL 환자의 생존 유무를 판단하기 위해 MGE 변수의 적정 cut-off point를 구하고자 한다. “Maxstat.test” 함수를 이용하여 생존형 결과 변수와 기간, 독립 변수를 넣어준 후, smethod는 “LogRank”로 지정한다. 그 결과, 표준화된 로그 순위 값은 3.18이며, 적정 cut-off point는 0.186임을 확인할 수 있으며, Fig. 3를 통해 MGE가 3.18일 때 표준화된 로그 순위 통계량이 가장 큰 것을 볼 수 있다. Modhukur 등[15]을 통해 생존형 결과 변수에서 로그 순위 통계량을 이용하여 cut-off point를 찾은 사례를 참고할 수 있다.

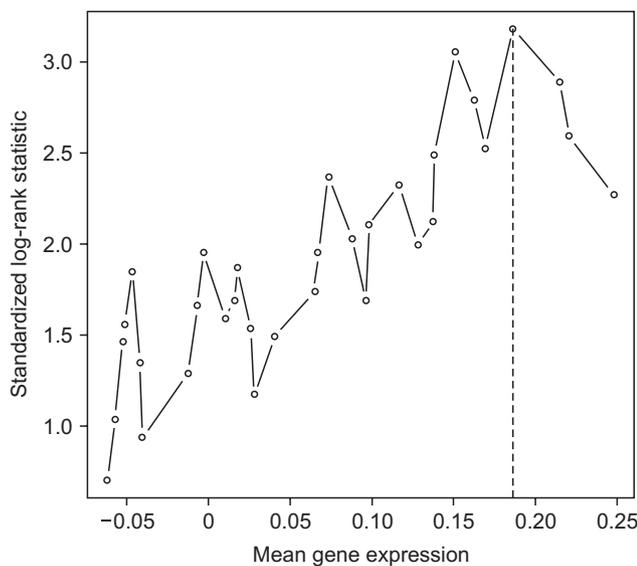


Fig. 3. Standardized log-rank statistic plot.

[R code]

```
>library(maxstat)
>library(survival)
#maxstat
>data(DLBCL)
>mstat<-maxstat.test(Surv(time,cens)~MGE,data
=DLBCL,smethod="LogRank",pmethod="exact-
Gauss",abseps=0.01)
```

**기타 방법**

결과 변수가 생존형 변수일 때 cut-off point를 찾는 방법으로는 조건부 추론 나무를 이용하는 방법이 있다. 이 방법은 “party” 패키지의 “ctree” 함수를 통해 구현 가능하다[16]. “Ctree” 함수는 결과 변수에 결측이 있으면 오류가 나기 때문에 cens 변수에 결측값을 제거한 후 시행해야 한다. “Ctree” 함수에 생존 함수를 넣은 후 돌려주면 최적의 cut-off point는 0.186으로, 이 예제에서는 로그 순위 통계량을 이용한 “maxstat” 패키지와 동일한 결과를 얻을 수 있다.

[R code]

```
>library(party)
>data=DLBCL[!is.na(DLBCL$cens),]
>out=ctree(Surv(time,cens)~MGE,data=data)
>out
```

**Cut-off Point 찾을 시 주의사항**

최적의 cut-off point는 가지고 있는 자료에서 결정이 되기 때문에 많은 표본 수를 확보하여 cut-off point를 끊어야 한다. 자료의 개수가 적을 경우, 끊어진 cut-off point는 과소 추정되거나 과대 추정될 수 있어 신뢰성이 떨어진다. 적절한 표본 수가 확보가 되었다면, 기본적으로 확인해야 할 것은 결과 변수의 형태이다. 먼저 끊고자 하는 cut-off point가 어떤 결과 변수를 위한 것인지, 그 결과 변수는 이분형인지, 순위형인지, 생존형인지 확인해야 할 것이다. 그렇게 결과 변수의 유형에 따라 어떤 방법을 선택하여 cut-off point를 끊을 것인지 정했다면, 두 번째로 확인해야 할 것은 끊고자 하는 변수의 크기가 커질수록 사건이 발생하는 것인지, 작을수록 사건이 발생하는 것인지 방향성을 확인하여야 한다. 특히 이분형 결과

**Table 6.** Method of finding optimal cut-off point according to outcome variables

Outcome	Method	R package
Binary	Maximize Youden index	pROC
Ordinal	Maximize Extended Youden index	DiagTest3gr
Survival	Maximize Log-rank statistics	maxstat

변수를 위하여 cut-off point를 끊을 때 방향을 생각하지 않고 ROC 곡선의 AUC를 구할 경우, 그 값이 낮게 계산될 수 있으며, 결과적으로 최적의 cut-off point를 찾는 것이 어려워진다. 세 번째는 cut-off point가 결정된 후, 이 수치가 실제로 나올 수 있는 값인지, 임상적으로 적절한 값인지 확인해야 할 것이다. 마지막으로 정해진 최적의 cut-off point는 검증을 통해 실제로 사용가능한 것인지 확인해야 한다. Cut-off point를 끊은 자료와 새롭게 구축한 자료에서 성능 평가를 하여 cut-off point의 성능이 적절한지 확인이 되어야 실제로 사용 가능한 cut-off point로 나아갈 수 있다. 성능 평가를 하는 방법으로는 이분형 자료에서는 민감도나 특이도 등이 있을 수 있으며, 순위형 자료에서는 se, sm, sp, 생존형 자료에서는 로그 순위 통계량 등을 이용하여 평가할 수 있다. 지금까지 설명한 cut-off point를 찾는 방법에 대한 정리 표는 Table 6과 같다.

## DISCUSSION

본 논문은 임상 연구에서 결과 변수가 이분형, 순위형 또는 생존 변수일 때 연속형 독립 변수의 cut-off point를 끊는 방법에 대해 알아보았다. 이분형 변수에서는 Youden index를 이용하여 그 값이 가장 커지는 곳을 최적의 cut-off point로 찾을 수 있으며, 순위형 변수에 대해서는 확장된 Youden index를 이용하여 찾을 수 있었다. 생존 변수에 대해서는 cut-off point를 기준으로 두 그룹으로 나누어 로그 순위 통계량을 계산하여 그 값이 가장 커지는 cut-off point를 최적의 cut-off point로 정하였다. 결과 변수에 따라 cut-off point를 결정하는 방법은 다르지만 기본적으로 특정 통계량을 최대화시키거나 판별 능력을 최대화하는 지점으로 결정하게 된다. 본 논문에서는 결과 변수에 따라 최적의 cut-off point를 설정할 수 있는 대표적인 몇 가지만 살펴보았다. 이외에도 다양한 cut-off point를 끊는 방법이 존재하기 때문에

연구자들은 자료의 특성, 질병의 특성에 따라 본 논문을 기반으로 적절한 방법을 찾아 끊기를 바란다. 적정 cut-off point를 끊는 것은 임상적으로 질병을 판단하거나 예방하는데 도움을 줄 수 있다. 하지만, cut-off point는 표본 자료를 기반으로 결정된 값이므로 항상 내부 자료와 외부 자료에 대한 성능 검증이 필요하며 결정된 값이 임상적으로 옳은지 판단해야 할 것이다. 본 논문을 기반으로 임상 연구자들이 연속적으로 측정된 바이오마커나 원인 인자에 대한 최적의 cut-off point를 찾아내어 올바른 의사결정을 할 수 있기를 바란다.

## FUNDING

None.

## CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

## AUTHOR'S CONTRIBUTIONS

Conceptualization: Hye Sun Lee. Supervision: Hye Sun Lee. Writing-original draft: Juyeon Yang. Writing-review&editing: Hye Sun Lee, Juyeon Yang.

## ORCID

Juyeon Yang, <https://orcid.org/0000-0002-7621-5150>

Hye Sun Lee, <https://orcid.org/0000-0001-6328-6948>

## REFERENCES

- Schisterman EF, Perkins NJ, Liu A, Bondell H. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology* 2005;16:73-81. <https://doi.org/10.1097/01.ede.0000147512.81966.ba>
- Ruopp MD, Perkins NJ, Whitcomb BW, Schisterman EF. Youden Index and optimal cut-point estimated from

- observations affected by a lower limit of detection. *Biom J* 2008;50:419-430. <https://doi.org/10.1002/bimj.200710415>
3. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. *Biom J* 2005;47:458-472. <https://doi.org/10.1002/bimj.200410135>
  4. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12:77. <https://doi.org/10.1186/1471-2105-12-77>
  5. Shin YG, Yoo J, Kwon HJ, et al. Histogram and gray level co-occurrence matrix on gray-scale ultrasound images for diagnosing lymphocytic thyroiditis. *Comput Biol Med* 2016;75:257-266. <https://doi.org/10.1016/j.compbiomed.2016.06.014>
  6. Nahm FS. Receiver operating characteristic curve: overview and practical use for clinicians. *Korean J Anesthesiol* 2022;75:25-36. <https://doi.org/10.4097/kja.21209>
  7. Johnson NP. Advantages to transforming the receiver operating characteristic (ROC) curve into likelihood ratio co-ordinates. *Stat Med* 2004;23:2257-2266. <https://doi.org/10.1002/sim.1835>
  8. Xiong C, van Belle G, Miller JP, Morris JC. Measuring and estimating diagnostic accuracy when there are three ordinal diagnostic groups. *Stat Med* 2006;25:1251-1273. <https://doi.org/10.1002/sim.2433>
  9. Luo J, Xiong C. Youden Index and associated cut-points for three ordinal diagnostic groups. *Commun Stat Simul Comput* 2013;42:1213-1234. <https://doi.org/10.1080/03610918.2012.661906>
  10. Luo J, Xiong C. DiagTest3Grp: an R package for analyzing diagnostic tests with three ordinal groups. *J Stat Softw* 2012;51:1-24. <https://doi.org/10.18637/jss.v051.i03>
  11. Lausen B, Schumacher M. Maximally selected rank statistics. *Biometrics* 1992;48:73-85. <https://doi.org/10.2307/2532740>
  12. Hothorn T, Lausen B. On the exact distribution of maximally selected rank statistics. *Comput Stat Data Anal* 2003;43:121-137. [https://doi.org/10.1016/S0167-9473\(02\)00225-6](https://doi.org/10.1016/S0167-9473(02)00225-6)
  13. Bland JM, Altman DG. The logrank test. *BMJ* 2004;328:1073. <https://doi.org/10.1136/bmj.328.7447.1073>
  14. Hothorn T, Hothorn MT, Suggests TH. Package ‘maxstat’. 2017. <https://cran.rproject.org/web/packages/maxstat/maxstat.pdf>. (accessed October 19, 2022).
  15. Modhukur V, Iljasenko T, Metsalu T, Lökk K, Laisk-Podar T, Vilo J. MethSurv: a web tool to perform multivariable survival analysis using DNA methylation data. *Epigenomics* 2018;10:277-288. <https://doi.org/10.2217/epi-2017-0118>
  16. Hothorn T, Hornik K, Strobl C, Zeileis A, Hothorn MT. Package ‘party’. *Packag Ref Man Party Version 0.9-998* 2015;16:37.