



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](#)

생존 자료에서 새로운 인자의  
예측력을 평가하는 방법의 비교



연세대학교 대학원  
의학전산통계학협동과정

의학통계학전공

추 순 규

생존 자료에서 새로운 인자의  
예측력을 평가하는 방법의 비교

지도교수 정 인 경

이 논문을 석사 학위논문으로 제출함



2015년 12월 일

연세대학교 대학원

의학전산통계학협동과정

의학통계학전공

추 순 규

# 추순규의 석사 학위논문을 인준함

심사위원 정 인 경 인

심사위원 남 정 모 인

심사위원 김 윤 남 인



연세대학교 대학원

2015년 12월 일

## 감사의 글

2015년과 석사과정의 함께 마무리 되어가는 시점에서 본 논문을 완성하기까지 밀어주고 끌어주신 많은 분들께 감사의 마음을 전하고자 합니다.

먼저 많이 부족한 학생임에도 불구하고 지도교수님으로서 아낌없는 지도와 격려로 나아갈 길을 찾아 주셨던 정인경 교수님께 감사드립니다. 항상 연구하는 모습을 보여주시며 졸업 후에도 계속해서 공부를 해야겠다는 결심을 갖게 해 주셨습니다. 교수님께서 보여주신 모습처럼 앞으로도 끊임없이 연구하며 살아가겠습니다. 항상 열정적으로 학업과 인생을 가르쳐 주신 남정모 교수님께 감사드립니다. 넓은 시야로 세상을 보라는 말씀 평생 잊지 않겠습니다. 가장 가까이에서 조언을 해주시고 폭넓은 지식을 전해주시는 송기준 교수님께 감사드립니다. 전공지식 뿐만 아니라 꼭 알아야 될 상식 등은 제가 살아가는 동안 큰 도움이 될 것입니다. 본 논문을 작성하는 동안 많은 도움을 주신 김윤남 선생님께 감사드립니다. 꼼꼼하게 논문의 개선사항을 체크해 주신 덕분에 무사히 논문을 작성할 수 있었습니다.

2년간의 석사 생활을 하는 동안 참 고마운 선배님과 동기, 후배들이 있었습니다. 석사 학위 논문의 주제를 고민하고 있을 때 도움을 주고 틀을 잡아 주신 이혜선 선배, 종종 학과에 찾아와 고민을 들어주고 웃음을 주었던 손낙훈 선배, 가끔 스트레스를 날려버릴 수 있도록 웃음을 주고 활력을 불어넣어준 노윤호 선배, 많은 것을 알려주고 도움을 준 조혜수 선배, 1년 동안 같이 생활하며 즐겁게 공부할 수 있도록 응원해 준 조호진 선배와 주우현 선배, 한 학기 선배로서 항상 좋은 말만 해준 최효진 선배와 이지현 선배, 석사 생활동안 서로의 힘이 되어준 성준형과 성환형, 항상 해야 될 일을 챙겨주고 도와준 세희와 세영, 챙겨주지 못했음에도 항상 잘 따라준 지유, 세정, 소정, 모든 열심히 나서서 열심히 해준 명균, 경민, 지윤, 덕성, 가장 힘든 막내의 역할을 너무나도 잘 해준 해린, 지현, 모두에게 감사합니다.

초등학생 때부터 언제나 옆에서 힘이 되어준 종대, 범준, 자주는 못 만나지만 가끔 만날 때 어제도 만난 사이 같은 민호, 호슴, 경창, 짓지 너무 고맙다.

하고 싶은 일이 생겼을 때 아무 말 없이 적극적으로 지원해 준 어머니, 아버지 정말 감사하고 사랑합니다. 마지막으로 미경아 항상 내 옆에 있어줘서 고맙고 사랑한다.



# 차 례

표 차례	iv
그림 차례	iv
국문 요약	vi
제1장 서론	1
1.1. 연구 배경 및 목적	1
1.2. 연구 내용 및 방법	2
1.3. 논문의 구성	2
제2장 이론적 배경	3
2.1. 이분형 종속변수	3
2.1.1. ROC curve & AUC	3
2.1.2. NRI	5
2.1.3. Continuous NRI	8
2.2. 생존 자료	9
2.2.1. Cox 의 비례위험모형	9
2.2.2. Harrell의 c-index	10
2.2.3. Heagerty의 integrated AUC	12
2.2.4. Uno의 c-index	14
2.2.5. Pencina와 Uno 의 NRI	15
제3장 모의실험	17
3.1. 모의실험 설계	17
3.1.1. 생존 시간의 생성	17
3.1.2. 중도절단 시간의 생성	18
3.1.3. 모의실험 설계	18

3.2. 모의실험 결과	19
3.2.1 중도절단 형태가 랜덤한 경우	19
3.2.2 일정 시점 이후 모든 대상이 중도절단 된 경우	20
3.2.3 중도절단 시간이 공변량과 독립이 아닌 경우	20
제4장 실제 자료 적용	27
제5장 결론	29
참고 문헌	31
부록	33
A-1. 중도절단 형태가 랜덤하고, 중도절단 비율이 0인 경우의 각 방법에 대한 추정치 및 추정치의 표준오차, 제 1종 오 류, 검정력	34
A-2. 중도절단 형태가 랜덤하고, 중도절단 비율이 0.3인 경우의 각 방법에 대한 추정치 및 추정치의 표준오차, 제 1종 오 류, 검정력	35
A-3. 중도절단 형태가 랜덤하고, 중도절단 비율이 0.5인 경우의 각 방법에 대한 추정치 및 추정치의 표준오차, 제 1종 오 류, 검정력	36
A-4. 중도절단 형태가 랜덤하고, 중도절단 비율이 0.7인 경우의 각 방법에 대한 추정치 및 추정치의 표준오차, 제 1종 오 류, 검정력	37
B-1. 시점 $\tau$ 에서 모든 대상이 중도절단 되고, 중도절단 비율이 0인 경우의 각 방법에 대한 추정치 및 추정치의 표준오차, 제 1종 오류, 검정력	38
B-2. 시점 $\tau$ 에서 모든 대상이 중도절단 되고, 중도절단 비율이 0.3인 경우의 각 방법에 대한 추정치 및 추정치의 표준오 차, 제 1종 오류, 검정력	39



B-3. 시점 $\tau$ 에서 모든 대상이 중도절단 되고, 중도절단 비율이 0.5인 경우의 각 방법에 대한 추정치 및 추정치의 표준오차, 제 1종 오류, 검정력	· · · · · 40
B-4. 시점 $\tau$ 에서 모든 대상이 중도절단 되고, 중도절단 비율이 0.7인 경우의 각 방법에 대한 추정치 및 추정치의 표준오차, 제 1종 오류, 검정력	· · · · · 41
C-1. 중도절단 시간이 생존 시간 $T$ 와는 독립이면서 공변량과는 독립이 아닐 때, 중도절단 비율이 0인 경우의 각 방법에 대한 추정치 및 추정치의 표준오차, 제 1종 오류, 검정력	· · · · · 42
C-2. 중도절단 시간이 생존 시간 $T$ 와는 독립이면서 공변량과는 독립이 아닐 때, 중도절단 비율이 0.3인 경우의 각 방법에 대한 추정치 및 추정치의 표준오차, 제 1종 오류, 검정력	· · · · · 43
C-3. 중도절단 시간이 생존 시간 $T$ 와는 독립이면서 공변량과는 독립이 아닐 때, 중도절단 비율이 0.5인 경우의 각 방법에 대한 추정치 및 추정치의 표준오차, 제 1종 오류, 검정력	· · · · · 44
C-4. 중도절단 시간이 생존 시간 $T$ 와는 독립이면서 공변량과는 독립이 아닐 때, 중도절단 비율이 0.7인 경우의 각 방법에 대한 추정치 및 추정치의 표준오차, 제 1종 오류, 검정력	· · · · · 45
영문 요약	· · · · · 46

## 표 차 례

표 1. 재분류표 . . . . .	6
표 2. 실제 자료에 대한 다섯 가지 통계량의 결과 . . . . .	28

## 그 립 차 례

그림 1. 중도절단 형태가 랜덤하고 대상의 수가 50인 경우의 제 1종 오류와 검정력 . . . . .	22
그림 2. 중도절단 형태가 랜덤하고 대상의 수가 100인 경우의 제 1종 오류와 검정력 . . . . .	22
그림 3. 중도절단 형태가 랜덤하고 대상의 수가 200인 경우의 제 1종 오류와 검정력 . . . . .	23
그림 4. 시점 $\tau$ 에서 모든 대상이 중도절단 되고 대상의 수가 50인 경우의 제 1종 오류와 검정력 . . . . .	23
그림 5. 시점 $\tau$ 에서 모든 대상이 중도절단 되고 대상의 수가 100인 경우의 제 1종 오류와 검정력 . . . . .	24
그림 6. 시점 $\tau$ 에서 모든 대상이 중도절단 되고 대상의 수가 200인 경우의 제 1종 오류와 검정력 . . . . .	24
그림 7. 중도절단 시간이 생존 시간 $T$ 와는 독립이면서 공변량과는 독립이 아니고 대상의 수가 50인 경우의 제 1종 오류와 검정력 . . . . .	25
그림 8. 중도절단 시간이 생존 시간 $T$ 와는 독립이면서 공변량과는 독립이 아니고 대상의 수가 100인 경우의 제 1종 오류와 검정력 . . . . .	25
그림 9. 중도절단 시간이 생존 시간 $T$ 와는 독립이면서 공변량과는 독립이 아니고 대상의 수가 200인 경우의 제 1종 오류와 검정력 . . . . .	26

## 국 문 요 약

### 생존 자료에서 새로운 인자의 예측력을 평가하는 방법의 비교

임상 분야에서 질병의 유무를 예측하는 다양한 통계적 방법이 존재한다. 특히, 기존의 예측모형에 새로운 인자가 추가 되었을 때의 예측력 향상을 평가하기 위해  $c$ -index와 net reclassification index(NRI)를 이용한다. 예측력을 의미하는  $c$ -index를 각 모형에서 구하여 비교할 수 있고, 새로운 모형에서 질병의 유무를 더 잘 분류했는지를 NRI를 통해 판단할 수 있다. 하지만 종속변수가 이분형이 아닌 생존 자료일 때는 중도절단을 고려하여 평가해야 한다. 생존 자료에서 예측력 향상을 평가하는 방법으로 Harrell의  $c$ -index, Heagerty의 시간종속 ROC curve를 이용한  $iAUC$ (integrated Area Under the Curve), Uno의  $c$ -index, Pencina의 NRI, Uno의 NRI 등이 제안되었다. 기존 모형과 새로운 인자가 추가된 예측모형에서의  $c$ -index 또는  $iAUC$  값을 각각 구하여 그 차이를 통해 예측력을 비교할 수 있으며, 두 예측모형으로부터 NRI 값을 구하여 새로운 모형이 생존여부를 더 잘 분류하는지를 판단할 수 있다.

본 연구는 다양한 상황을 가정한 모의실험을 통해 생존 자료에서 예측력 향상을 평가하는 통계적인 방법들의 성능을 비교하였고, 실제 자료에 적용해 보았다. 질병의 유무를 예측하는 두 예측모형의 예측력을 비교하는 데 있어 더 민감하다고 알려진 NRI가 생존 자료의 특성에 맞게 변화되었을 때는 Harrell과 Uno의  $c$ -index, Heagerty의  $iAUC$ 에 비해 덜 민감한 결과를 나타냈다.

---

핵심되는 말 :  $c$ -index, 시간종속 ROC curve, NRI, 생존 분석, 예측모형 비교

# 제 1 장 서론

## 1.1 연구 배경 및 목적

임상 분야에서 질병 유무를 예측하는 예측모형에 대한 예측력을 평가하는 다양한 통계 방법이 존재한다. 일반적으로 질병 유무와 같은 이분형 종속변수에 대한 예측력 평가는 ROC curve (Receive Operating Characteristic curve)의 아래 면적(AUC; Area Under the Curve)을 이용하는 방법이 대표적이다. 단순히 하나의 예측모형에 대한 예측력 평가가 목적이 아니라 기존 예측모형보다 새로운 인자가 추가로 포함된 예측모형의 예측력이 더 뛰어난지 확인하는 것이 목적인 경우에는 두 모형의 AUC를 비교하여 확인할 수 있다. 기존 예측모형에 새로운 인자가 추가되었을 때, 기존 예측모형과 새로운 예측모형은 서로 독립이라고 가정할 수 없으므로 단순히 독립된 두 통계량을 비교하는 방법을 사용하는 것은 적절하지 않다는 연구결과가 있다(DeLong, DeLong, Clarke-Pearson 1988). DeLong 외(1988)은 독립이 아닌 두 모형의 AUC를 이용한 예측력 비교 방법을 제안하였으나 어떤 상황에서는 새로운 모형의 예측력이 더 뛰어나에도 불구하고 이를 판별하지 못하는 경우가 발생한다(Cook 2007). 이러한 문제를 보완하기 위해 Pencina 외(2008)는 AUC 대신 net reclassification improvement (NRI)와 integrated discrimination improvement (IDI)를 이용하여 예측력을 비교하는 방법을 제안하였다. NRI와 IDI는 두 예측모형에 대한 비교적 작은 예측력 차이에도 AUC 비교에 비해 민감하게 반응한다고 알려져 있다(Pencina 외 2008).

예측해야 하는 종속변수가 생존 자료일 경우, 앞에서 언급한 DeLong 외(1988)과 Pencina 외(2008)가 제안한 방법은 생존 자료가 가질 수 있는 정보를 고려하지 못하는 제약이 있다. 생존 자료에서 예측모형에 대한 예측력을 평가하는 방법으로 Harrell 외(1982)의 c-index, Heagerty 외(2005)의 시간종속 ROC curve를 이용한 *i* AUC, Uno 외(2011)의 c-index, Pencina 외(2011)의 NRI, Uno 외(2013)의 NRI가 대표

적이다. 종속변수가 생존 자료일 때, 생존 자료의 특성을 고려한 이런 방법들에 대하여 비교 연구된 바가 없다. 따라서 본 연구에서는 실제 임상에서 존재할 수 있는 여러 상황을 가정하여 위에 언급된 다섯 가지 방법을 비교해 보고자 한다.

## 1.2 연구 내용 및 방법

생존 자료에서 기존 예측모형에 새로운 인자가 추가 되었을 때의 예측력 향상을 평가하는 방법으로 사용되는 Harrell 외(1982)의 c-index, Heagerty 외(2005)의 시간종속 ROC curve를 이용한 iAUC, Uno 외(2011)의 c-index, Pencina 외(2011)의 NRI, Uno 외(2013)의 NRI 방법에 대해 이분형 종속변수일 경우를 포함하여 문헌검토를 하였다. 중도절단의 종류, 비율, 표본의 수, 새로운 인자의 유의성 등 여러 상황을 가정한 생존 자료를 생성하여 모의실험을 통해 변별력을 살펴보고자 한다. 또한, 실제 자료에 적용되었을 경우에는 어떤 결과가 도출되는지 살펴보고자 한다.

## 1.3 논문의 구성

제 1장에서는 연구의 배경과 목적, 내용 및 방법을 소개한다. 2장에서는 종속변수가 이분형일 경우와 생존 자료일 경우를 나누어 예측모형의 예측력을 구하는 방법과 두 예측모형의 예측력을 비교하는 방법을 살펴본다. 3장에서는 모의실험을 통해 생존 자료에서 기존 예측모형에 새로운 인자가 추가 되었을 때의 예측력 향상 여부를 다섯 가지 방법을 통해 평가하고 비교한다. 4장에서는 다섯 가지 방법을 실제 자료에 적용하였을 때의 결과를 비교한다. 마지막으로 5장에서는 결론 및 고찰에 대해 논의한다.

## 제 2 장 이론적 배경

2.1에서는 종속변수가 이분형 자료인 경우에 대해서 다루고자 한다. 단일 예측모형의 예측력을 평가하는 통계적인 방법에 대한 이론을 언급하고, 두 예측모형의 예측력을 비교하는 방법에 대해 소개한다. 2.2에서는 종속변수가 생존 자료인 경우에 대해 다룬다. 단일 예측모형의 예측력을 평가하는 통계적인 방법에 대한 이론을 소개한다. 또한 두 예측모형의 예측력을 비교하는 방법에 대해 소개한다.

### 2.1 이분형 종속변수

2.1.1에서는 예측모형의 예측력을 판단하는 방법인 ROC curve에 대해 소개한다. 또한 독립된 두 예측모형과 독립이 아닌 두 예측모형의 ROC curve의 아래 면적인 AUC를 비교하는 방법에 대해 소개한다. 2.1.2에서는 독립이 아닌 두 예측모형의 예측력을 판단하는 방법인 NRI에 대해 소개한다. 2.1.3에서는 예측확률의 범주를 나누지 않고 연속적으로 계산하는 방법인 cNRI에 대해 소개한다.

#### 2.1.1 ROC curve & AUC

임상 분야에서 질병의 유무를 의미하는 이분형 변수가 종속변수로 사용될 때, 질병의 유무에 대한 판단이 주된 목적인 경우가 있다. 질병의 유무를 적은 비용으로 더 정확하게 판단하기 위해 다양한 예측모형을 구축하고 비교하는 것이 좋다. 이 때, 가장 흔히 사용되는 통계적인 방법으로는 ROC curve가 있다. ROC curve는 민감도(sensitivity)와 특이도(specificity)를 이용해 구할 수 있다. 실제 질병이 있는 대상을 검사 결과 질병이 있다고 판단할 확률을 민감도라고 하며, 실제 질병이 없는 대상을 검사 결과 질병이 없다고 판단할 확률을 특이도라고 한다. 예측모형을 통해 얻어진 질병이 있을 확률을  $p$ 라고 했을 때, 보통은  $p$ 가 0.5보다 클 때 질병이 있다고

판단한다. 그러나 이는 상황에 따라 달라질 수 있으며 어떤 상황에서는  $p$ 가 0.1보다 클 때 질병이 있다고 판단할 수도 있다. 이와 같은 상황은 특이도보다 민감도가 더 중요한 상황에서 발생할 수 있다. ROC curve는 질병이 있다고 판단할 확률을 이동하면서 민감도와 특이도를 이용하여 구할 수 있다. 예를 들어 예측확률  $p$ 가 0.1을 기준으로 질병의 유무를 판단한다고 하면, 그 때의 민감도와 특이도를 구하여  $y$  축을 민감도,  $x$  축을 (1-특이도)로 하는 그래프에 한 점을 찍는다. 질병이 있다고 판단할  $p$ 의 기준을 0에서 1까지 이동하면서 민감도와 특이도를 구하여 그래프에 점을 찍으면 ROC curve가 완성된다. 이렇게 구해진 ROC curve의 아래 면적을 AUC라고 하며, 0.5에서 1 사이 값을 갖는 예측모형의 예측력이라고 정의한다. AUC는 1에 가까울수록 예측모형의 예측력이 뛰어나다고 하며, 0.5에 가까울수록 예측력이 떨어진다고 판단한다.

예측모형이 하나가 아닌 두 개인 경우, 각 예측모형의 AUC가 1에 더 가까운 모형이 예측력이 뛰어나다고 할 수 있다. 또한, 이러한 두 예측모형의 예측력 차이가 통계적으로 유의한 차이인지를 평가해야 한다. 첫 번째 예측모형을  $Model_1$ , 두 번째 예측모형을  $Model_2$ 라고 하고 각 모형에서 얻어진 예측력을  $AUC_1$ ,  $AUC_2$ 라고 하자. 직관적으로 두 통계량의 차이를 두 통계량의 차이에 대한 표준오차로 나누어 주면 통계적인 판단을 할 수 있다(Hanley, McNeil 1983).

$$\frac{AUC_1 - AUC_2}{SE(AUC_1 - AUC_2)} \sim Z(0,1)$$

두 예측모형이 독립일 경우에는  $SE(AUC_1 - AUC_2) = \sqrt{SE^2(AUC_1) + SE^2(AUC_2)}$ 로 계산할 수 있다. 그러나 두 예측모형이 독립이 아닐 경우에는 표준오차에 대한 식이 아래와 같이 정의된다.

$$SE(AUC_1 - AUC_2) = \sqrt{SE^2(AUC_1) + SE^2(AUC_2) - 2rSE(AUC_1)SE(AUC_2)}$$

위 식에서  $r$ 은 두 AUC의 상관성을 의미한다. 두 예측모형이 독립일 경우에는 예측력

의 차이에 대한 검정통계량을 간단히 구할 수 있다. 그러나 두 예측모형이 독립이 아닐 경우에는 예측력 차이에 대한 표준오차를 구하기 어렵다. 본 연구에서는 두 예측모형이 독립이 아닌 경우를 다루며, Efron (1979)이 제시한 붓스트랩(bootstrap) 방법을 통해 두 예측모형의 예측력을 비교 분석한다.

## 2.1.2 NRI

2.1.1과 2.1.2에서는 두 예측모형에서의 ROC curve에 대한 AUC를 각각 구하고 두 AUC의 차이를 통계적으로 검정하는 방법에 대해 살펴보았다. 2.1.3에서는 앞에서 다룬 방법과는 달리 두 예측모형을 기반으로 분류표를 만들어 어떤 모형이 더 정확하게 질병의 유무를 분류해 내는지 비교하는 NRI 방법에 대해 살펴볼 것이다.

NRI는 새로운 인자가 기존 예측모형에 추가되었을 때 예측력 향상을 평가하기 위한 목적으로 Pencina 외(2008)에 의해 제안되었다. 따라서 이 방법은 하나의 예측모형이 다른 하나의 예측모형에 내재(nested)되어 있는 경우, 즉, 두 예측모형이 종속되어 있는 경우에 대해서 예측력을 비교할 때 사용된다. 기존 예측모형과 새로운 인자가 포함된 예측모형을 각각  $Model_1$ ,  $Model_2$ 라고 하면, 각 예측모형에 대한 로지스틱 회귀분석 모형식은 다음과 같다.

$$Model_1 : \log\left(\frac{p_1(Y=1)}{1-p_1(Y=1)}\right) = \alpha_1 + \beta_{1,1}x_1 + \dots + \beta_{n,1}x_n$$

$$Model_2 : \log\left(\frac{p_2(Y=1)}{1-p_2(Y=1)}\right) = \alpha_2 + \beta_{1,2}x_1 + \dots + \beta_{n,2}x_n + \beta_{n+1,2}x_{n+1}$$

$Y=1$ 인 경우를 질병이 있는 집단(event)이라고 할 때,  $Model_1$ ,  $Model_2$ 를 통해 event 집단이 될 예측확률  $p_1$ ,  $p_2$ 를 갖는다.



$$p_1(Y=1) = \frac{1}{1 + \exp\{-(\alpha_1 + \beta_{1,1}x_1 + \dots + \beta_{n,1}x_n)\}}$$

$$p_2(Y=1) = \frac{1}{1 + \exp\{-(\alpha_2 + \beta_{1,2}x_1 + \dots + \beta_{n,2}x_n + \beta_{n+1,2}x_{n+1})\}}$$

각  $i$  번째 대상은 모형에서의 예측확률을 통해 event 집단에 속할 확률을 가지며, 어떤 cut point를 기준으로 event 집단에 속할 확률을 낮음, 중간, 높음으로 구분할 수 있다.  $Model_k$ ,  $k=1,2$ 에서 구해진 예측확률  $p_k$ 를 cut point를 기준으로 분류표를 작성하면 다음과 같다.

표 1.은 cut point를 기준으로 하여 범주를 총 세 지점으로 잡은 경우의 재분류표를 나타낸다. NRI는 일반적으로 cut point를 기준으로 범주를 두 개 혹은 세 개로 나누어 이용한다. 재분류표를 이용하여 기존 예측모형에 비해 새로운 인자가 포함된 예측모형이 event 집단과 non-event 집단을 더 잘 분류했는지를 판단한다. 분류표를 이용한 NRI의 정의는 다음과 같다.

$$NRI = [P(up|Y=1) - P(down|Y=1)] - [P(up|Y=0) - P(down|Y=0)]$$

표 1. 재분류표

Event	$Model_2$			Non-event	$Model_2$		
	낮음	중간	높음		낮음	중간	높음
$Model_1$ 낮음	a	b	c	$Model_1$ 낮음	j	k	l
$Model_1$ 중간	d	e	f	$Model_1$ 중간	m	n	o
$Model_1$ 높음	g	h	i	$Model_1$ 높음	p	q	r

$P(up|Y=1)$ 는 event 집단에서  $Model_2$ 의 예측확률이 더 높게 분류된 확률을 의미한다. 반대로  $P(down|Y=1)$ 는 event 집단에서  $Model_2$ 의 예측확률이 더 낮게 분류된 확률을 의미한다.

$$\begin{aligned}\hat{P}(up|Y=1) &= \frac{b+c+f}{a+b+c+d+e+f+g+h+i} \\ \hat{P}(down|Y=1) &= \frac{d+g+h}{a+b+c+d+e+f+g+h+i} \\ \hat{P}(up|Y=0) &= \frac{k+l+o}{j+k+l+m+n+o+p+q+r} \\ \hat{P}(down|Y=0) &= \frac{m+p+q}{j+k+l+m+n+o+p+q+r}\end{aligned}$$

추정된 확률들을 이용하여 다음과 같이 NRI를 나타낼 수할 수 있다.

$$\begin{aligned}\widehat{NRI} &= (\hat{P}(up|Y=1) - \hat{P}(down|Y=1)) + (\hat{P}(down|Y=0) - \hat{P}(up|Y=0)) \\ &= \left[ \left( \frac{b+c+f}{n_{event}} \right) - \left( \frac{d+g+h}{n_{event}} \right) \right] + \left[ \left( \frac{m+p+q}{n_{nonevent}} \right) - \left( \frac{k+l+o}{n_{nonevent}} \right) \right]\end{aligned}$$

새로운 예측모형의 예측력이 더 뛰어나다면 NRI 값은 0보다 커질 것이다. NRI 값이 0보다 큰 값을 갖는지 검정하기 위한 검정통계량은 다음과 같다(Pencina 외 2008).

$$z = \frac{\widehat{NRI}}{\widehat{SE}(\widehat{NRI})}$$

$$\text{where } \widehat{SE}(\widehat{NRI}) = \sqrt{\frac{\hat{P}(down|Y=0) + \hat{P}(up|Y=0)}{n_{nonevent}} + \frac{\hat{P}(up|Y=1) + \hat{P}(down|Y=1)}{n_{event}}}$$

### 2.1.3 Continuous NRI

기존의 NRI는 cut point를 기준으로 범주를 세 개 또는 두 개를 사용한다(Pencina 외 2008). 다른 연구에서는 범주를 네 개로 나누거나 범주를 나누지 않고 적용한 경우도 있었다. 그러나 절대적인 범주의 수가 존재한다고는 할 수 없으며, 범주를 나눌 때의 cut point가 논쟁이 되기도 하였다. 이런 문제를 해결하기 위해 Pencina 외 (2011)이 범주를 나누지 않는 방법을 제안하였다. 이 방법은 기존 예측모형보다 새로운 예측모형의 예측확률이 높으면 up, 낮으면 down으로 분류한다. 범주를 나누지 않기 때문에 continuous NRI 혹은 category-free NRI라고 이름을 붙였으며, cNRI라고 간략하게 표기한다. cNRI는 다음과 같이 정의된다.

$$cNRI = E\{sign(p_2 - p_1)|Y = 1\} + E\{-sign(p_2 - p_1)|Y = 0\}$$

$p_1, p_2$ 는 각각 기존 예측모형과 새로운 예측모형으로부터 얻어진 예측확률을 의미한다. event 집단과 non-event 집단의 독립을 가정하면, cNRI의 검정통계량은 다음과 같이 나타낼 수 있다.

$$z_{cNRI} = \frac{c\widehat{NRI}}{4 \left[ \frac{\widehat{P}(up|event)(1 - \widehat{P}(up,event))}{n_{event}} + \frac{\widehat{P}(up|nonevent)(1 - \widehat{P}(up,nonevent))}{n_{nonevent}} \right]}$$

검정통계량을 통해 유의한 결과를 얻게 되면, 새로운 예측모형과 기존 예측모형의 예측력에 유의한 차이가 있다고 할 수 있다.

## 2.2 생존 자료

이번 절에서는 생존 자료일 경우 예측모형의 예측력을 구하는 방법과 두 예측모형의 예측력을 비교하는 방법을 살펴본다. 2.2.1에서는 본 연구에서 사용 될 통계량들이 Cox의 비례위험모형을 가정하므로 Cox의 비례위험모형을 소개한다. 2.2.2에서는 예측모형의 예측력을 구하는 비모수적인 방법인 c-index를 생존 자료에 적용한 방법인 Harrell의 c-index에 대해 소개한다. 2.2.3에서는 시간에 따라 변하는 ROC curve의 아래 면적을 모든 시간  $t$ 에 대해 구하고 다시 그 아래 면적을 구하는 Heagerty의  $i$ AUC에 대해 소개한다. 2.2.4에서는 중도절단 분포에 의존하지 않는 Uno가 제안한 c-index에 대해 소개한다. 2.2.5에서는 NRI를 생존 자료에 적용하는 방법과 Pencina가 제안한 생존 자료에서의 cNRI, Uno가 제안한 거리를 이용한 NRI에 대해 소개한다.

### 2.2.1 Cox의 비례위험모형(Cox's proportional hazard model)

위험함수(hazard function)란 임의의  $t$ 시점까지 생존한 대상의 순간위험률이며, 수식적으로 표현하면 다음과 같다.

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}$$

위 식에서  $t$ 는 생존 시간을 의미하며, 위험함수는 생존함수(survival function)와 확률밀도함수(probability density function)로 표현된다. 확률밀도함수와 생존함수의 관계  $f(t) = -\frac{d}{dt}S(t)$ 를 이용하면 생존함수만으로 위험함수를 표현할 수 있다.

$$h(t) = -\frac{S'(t)}{S(t)} = \frac{\frac{d}{dt}S(t)}{S(t)} = -\frac{d}{dt}\log S(t)$$

Cox의 비례위험모형은 생존 시간과 공변량의 관계를 설명하기 위해 위험함수를 기초로 하는 회귀모형이다. Cox의 비례위험모형은 시간  $t$ 가 변화함에 따른 공변량에 대한 위험함수의 비  $\frac{h(t|X=x+1)}{h(t|X=x)}$ 는 일정하다는 가정을 만족해야 한다. Cox의 비례위험모형은 다음과 같다.

$$h(t|x) = h_0(t) \times \exp(\beta'x)$$

위 식에서  $h_0(t)$ 는 기저위험함수(baseline hazard function)이며,  $x$ 와  $\beta'$ 는 공변량 벡터와 해당 모수 열벡터를 의미한다. 본 연구는 Cox의 비례위험모형을 가정하여 예측력 향상을 평가하는 다양한 방법을 비교한다.

### 2.2.2 Harrell의 c-index

생존 자료에서 예측모형의 예측력을 구하는 대표적인 비모수적 방법으로 Harrell의 c-index가 있다. Harrell의 c-index는 이분형 종속변수에서 Hanley 외(1982)의 c-index를 구하는 비모수적인 방법을 생존 자료에 적용시킨 방법으로 중도절단을 고려하여 제시된 방법이다. 생존 시간을  $T_i$ , 중도절단 시간을  $C_i$ 라고 정의하면 관찰된 생존 시간은  $X_i = \min(T_i, C_i)$ 로 정의할 수 있다. 이때, 중도절단여부는 다음과 같이 정의 한다.

$$\Delta_i = \begin{cases} 1, & \text{if } X_i = T_i \\ 0, & \text{if } X_i \neq T_i \end{cases}$$

Harrell의 c-index를 정의하기 위해서는 비교 가능한 쌍과, 비교 불가능한 쌍을 정의해야 한다.  $i$  번째 대상과  $j$  번째 대상이 모두 중도절단된 경우이거나, 모두 중도절단 되지 않은 경우에는 생존 시간의 비교가 가능하다. 그러나  $i$  번째 대상과  $j$  번째

째 대상 중 하나의 대상만 중도절단이 된 경우는 두 가지 경우로 나뉘게 된다.

$$X_i < X_j, \Delta_i = 0 \text{ or } X_i > X_j, \Delta_j = 0,$$

이 경우 생존 시간의 비교가 불가능하나, 다른 하나의 경우는 생존 시간의 비교가 가능하며 그 형태는 다음과 같다.

$$X_i < X_j, \Delta_i = 1 \text{ or } X_i > X_j, \Delta_j = 1$$

모든 경우의 쌍 중에서 첫 번째 경우를 제외하면 모든 쌍이 비교 가능한 쌍이 된다. 비교 가능한 쌍에서  $X_i > X_j, \delta_j = 1$  일 경우 점수를 1 로,  $X_i < X_j, \delta_i = 1$  일 경우 점수를 0 으로 정의할 때, Harrell의 c-index는 다음과 같이 정의 된다(Harrell 외 1982).

$$C_{harrell} = \frac{\text{비교 가능한 쌍에서의 점수의 합}}{\text{비교 가능한 쌍의 수}}$$

만약 중도절단 된 대상이 없다면, 모든 쌍들이 비교 가능한 쌍이 되며 Mann-Whitney 통계량과 같아지게 된다.

기존 예측모형의 c-index와 새로운 인자가 추가로 포함된 예측모형의 c-index를 각각 구하여, 새로운 모형에서의 예측력이 더 향상 되었는지 판단하기 위해서는 두 c-index의 차이가 0 인지를 검정하면 된다. 검정을 위해서는 두 모형에서 얻어진 c-index의 차이에 대한 분산을 구하여 t 통계량을 이용할 수 있다. 그러나 두 모형을 독립이라고 간주할 수 없기 때문에 두 c-index의 차이에 대한 분산을 구하기 쉽지 않다. 그러므로 붓스트랩 방법을 이용하여 두 c-index의 차이에 대한 신뢰구간을 구하여 95% 신뢰구간에 0 이 포함되는지 여부로 예측력 향상을 판단할 수 있다.

### 2.2.3 Heagerty의 integrated AUC

Heagerty 외(2005)는  $x$  축을 시점  $t$ ,  $y$  축을 시점  $t$ 에서의 ROC curve의 아래 면적인 AUC 값으로 갖는 그래프를 그려 그 아래 면적을 적분하여  $iAUC$ (integrated AUC) 값을 구하였다. Heagerty 외(2005)는 총 세 가지의 시간종속 ROC curve를 제안하였는데 그 중 Incident/dynamic ROC curve를 본 연구에서 비교에 사용하였다. ROC curve는 모든 가능한 지점에서 민감도와 특이도를 표현한 곡선이다. 어떤 시점  $t$ 에서의 민감도와 특이도는 다음과 같이 정의된다.

$$sensitivity^I(c,t) : P(M_i > c | T_i = t) = P(M_i > c | dN_i^*(t) = 1)$$

$$specificity^D(c,t) : P(M_i \leq c | T_i > t) = P(M_i \leq c | dN_i^*(t) = 0)$$

위 식에서  $M_i$ 는 공변량 벡터와 회귀계수 벡터의 곱인  $\beta'x$ 를 의미한다. 또한  $dN_i^*(t) = I(T_i \leq t) - I(T_i \leq t-)$ 를 나타내며, 시점  $t$ 에서 사건이 발생한 경우 1, 아닌 경우 0을 의미한다.

시간  $t$ 에 따른 ROC curve를 시간종속 ROC curve 라 하며, Incident/dynamic ROC curve를  $ROC_t^{I/D}(p)$ 의 함수식으로 정의하였다.  $ROC_t^{I/D}(p)$ 의 정의는 아래와 같으며,

$$ROC_t^{I/D}(p) = TP_t^I\{[FP_t^D]^{-1}(p)\}$$

$p$ 는 dynamic false-positive rate,  $ROC_t^{I/D}(p)$ 는 incident true-positive rate을 의미한다. 이 때, 시점  $t$ 에서의 Incident/dynamic ROC curve에 대한 AUC는 다음과 같이 구할 수 있다.

$$AUC(t) = \int_0^1 ROC_t^{I/D}(p) dp$$

임의의 시점  $\tau$ 에서의 예측력은  $\tau$ 까지의 AUC를 적분하여 구할 수 있으며, 이를  $iAUC$ 로 정의한다. 그 식은 다음과 같다.

$$C^\tau = \int_0^\tau AUC(t) \cdot w^\tau(t) dt$$

위 식에서  $w^\tau(t)$ 는 가중치를 의미하며, 다음과 같이 정의된다.

$$w^\tau(t) = \frac{2 \cdot f(t) \cdot S(t)}{W^\tau}, \quad W^\tau = \int_0^\tau 2 \cdot f(t) \cdot S(t) dt = 1 - S^2(\tau)$$

본 연구에서는 Harrell의 c-index와 마찬가지로 기존의 예측모형과 새로운 인자가 포함된 예측모형에서 각각  $iAUC$ 를 구한 후 그 차이에 대한 신뢰구간을 붓스트랩 방법을 통해 구하여 다른 방법들과 성능을 비교하고자 한다.



## 2.2.4 Uno 의 c-index

Harrell 외(1982)이 제안한 생존 자료에서의 c-index는 모수  $\beta$ 가 자료의 중도절단 분포에 의존하여 추정되는 문제가 발생한다. 또한 생존 시간에 중도절단이 있는 경우 Harrell 외(1982)이 제안한 c-index와 Heagerty 외(2005)가 시점  $(0, \tau)$ 에서 제안한 보완된  $C_\tau$ 의 추정이 간단하지 않다(Uno 외 2011). Harrell의 c-index와 Heagerty의  $C_\tau$ 는 다음과 같이 표현할 수 있다.

$$C_{harrell} = P(M_i > M_j | T_i < T_j)$$

$$C_\tau = P(M_i > M_j | T_i < T_j, T_i < \tau)$$

위 식에서  $C_\tau$ 는 추적조사기간이  $(0, \tau)$ 일 경우를 나타낸다.

Uno(2009)는 생존 시간에 중도절단이 존재하는 경우에도 중도절단 분포에 무관한 c-index를 제안하였다. 제안한 비모수적 통계량은 아래와 같다.

$$\hat{C}_\tau = \frac{\sum_{i=1}^n \sum_{j=1}^n \Delta_i \hat{G}(X_i)^{-2} I(X_i < X_j, X_i < \tau) I(\beta' Z_i > \beta' Z_j)}{\sum_{i=1}^n \sum_{j=1}^n \Delta_i \hat{G}(X_i)^{-2} I(X_i < X_j, X_i < \tau)}$$

위 식에서  $I(\cdot)$ 는 지시함수를 의미하며,  $\hat{G}(\cdot)$ 는 중도절단 분포  $G(t) = P(D > t)$ 에 대한 Kaplan-Meier 추정치를 의미한다. 기존 예측모형과 새로운 인자가 포함된 예측모형에서 Uno 외(2011)가 제안한 c-index를 각각 구한 후, 붓스트랩 방법을 통해 두 모형에서의 c-index의 차이에 대한 95% 신뢰구간을 구하여 두 예측 모형의 예측력을 비교할 수 있다.

## 2.2.5 Pencina 와 Uno 의 NRI

2.2.2부터 2.2.4까지 c-index와 ROC curve를 기반으로 하는 통계량들을 살펴보았다. 위에서 언급한 세 가지 방법들은 작은 차이에 민감하지 않은 단점이 있다고 알려져 있다(Cook 2007). 이를 해결하기 위해 NRI, IDI 등의 통계량이 제안되었다. 2.2.5에서는 NRI를 생존 자료에 적용한 통계량을 살펴본다.

Uno 외(2013)는  $t_0$ 시점 이전에 event가 발생한 대상을 환자군으로 분류하고, 그렇지 않은 대상을 대조군으로 분류하여 NRI를 정의하였다. 기존 예측모형에서의 위험과 새로운 인자가 포함된 예측모형에서의 위험의 차이를 구하고 다음과 같이 거리를 의미하는 추정량을 제안하였다.

$$\hat{D}(\mathbf{x};t_0) = \hat{p}_2(\mathbf{x}_2;t_0) - \hat{p}_1(\mathbf{x}_1;t_0)$$

기존 예측모형과 새로운 예측모형의 공변량벡터를 각각  $\mathbf{x}_1$ ,  $\mathbf{x}_2$ 이라 하며,  $\hat{p}_1(\mathbf{x}_1;t_0)$ 는 첫 번째 예측모형에서 Cox의 비례위험모형을 통해 얻어진 위험 점수를 의미한다. 앞에서 정의된  $\hat{D}(\mathbf{x};t_0)$ 가  $t_0$  시점 이전에 event가 발생한 대상에서 양수 값을 가지게 되면, 새로운 예측모형의 예측력이 더 좋은 것을 의미한다. 반면,  $t_0$  시점 이전에 event가 발생하지 않은 대상에서는  $\hat{D}(\mathbf{x};t_0)$ 가 음수여야 새로운 예측모형의 예측력이 더 좋은 것을 의미한다. 앞에서 제안한 거리를 의미하는  $\hat{D}(\cdot |t_0)$ 를 이용하여 다음과 같은 두 함수를 정의할 수 있다.

$$F_n(s;t_0) = P\{\hat{D}(\mathbf{x};t_0) \leq s | T^0 \leq t_0\}$$

$$G_n(u;t_0) = P\{\hat{D}(\mathbf{x};t_0) \leq u | T^0 > t_0\}$$

$$\text{where, } (s, u) \in [-1, 1] \times [-1, 1]$$

위와 같이  $F_n$ 과  $G_n$ 을 정의한 후, 두 함수의 결합 함수의 그래프를  $x$  축을  $s$ ,  $y$  축을  $P(\hat{D} \leq s)$ 로 그릴 수 있다.  $T^0 \leq t_0$ 의 경우와  $T^0 > t_0$ 의 경우를 한 좌표평면 상에 그리면, Pencina 외(2011)와 Uno 외(2013)가 각각 제안한 cNRI와  $\text{NRI}_{Uno}$ 를 다음과 같이 정의할 수 있다. cNRI의 값은  $s$ 가 0인 지점에서의  $y$  축의 값 즉,  $P(\hat{D} \leq s)$ 의 차이를 의미한다.  $\text{NRI}_{Uno}$ 는  $P(\hat{D} \leq s)$  값이 0.5인 지점에서의  $x$  축의 값,  $s$ 의 차이를 의미한다.



## 제 3 장 모의실험

3 장에서는 2 장에서 소개한 다섯 가지 예측력 향상 평가 방법의 성능을 비교 평가한다. 실제 임상에서 존재할 만한 자료를 생성하여 각 상황에서 성능이 뛰어난 방법을 찾아내기 위해 모의실험을 수행한다.

### 3.1 모의실험 설계

#### 3.1.1 생존 시간의 생성

지수분포를 따르는 Cox의 비례위험모형에서 생존 시간을 생성하기 위해  $S_0(t) = e^{-\lambda t}$ 를 가정하였다.  $i$  번째 대상의 생존 시간은 다음과 같이 표현된다 (Bender, Augustin, Blettner 2005).

$$T_i = -\frac{\log(U_i)}{\lambda \exp(\beta' \mathbf{x})}$$

생존 시간을 생성할 때 독립변수  $\mathbf{x}$ 는 평균이 0 이고, 공분산이 0.1 인 다변량 정규분포를 가정하였다. 독립변수의 개수는 두 개이며, 두 번째 독립변수를 새로운 예측모형에 포함된 인자로 가정하였다. 여기서  $U_i$ 는 균등분포에서 생성된 (0,1)의 임의의 실수이고,  $\lambda$ 는 0.5를 사용하여 생존 시간을 생성하였다. 회귀계수  $\beta$ 는 다음과 같이 총 세 가지 상황을 가정하였다. 첫 번째 상황은  $\beta = (1, 0)$ 이다. 기존의 독립변수는 생존 시간에 영향을 미치며, 새로 추가된 인자는 생존 시간에 전혀 무관한 독립변수임을 가정하였다. 두 번째 상황은  $\beta = (1, 0.3)$ 이며, 새로 추가된 인자의 영향력을 다소 향상시켰다. 이 경우 Cox의 비례위험모형에서의 회귀 계수에 대한 유의확률은 유

의수준 0.05보다 크게 나타난다. 마지막 상황은  $\beta=(1, 0.5)$  으로 새로 추가된 인자의 영향력을 보다 향상시켰다. 마지막 상황의 경우 Cox의 비례위험모형에서의 새로 추가된 변수에 대한 회귀 계수의 유의확률은 유의수준 0.05 보다 굉장히 작게 나타난다.

### 3.1.2 중도절단 시간의 생성

중도절단 시간의 생성은 세 가지 상황을 가정한다. 첫 번째 상황은 생존 시간과 독립변수에 독립인 중도절단 시간을 생성하는 것이다. 두 번째 상황은 일정  $t$ 시점 이후의 생존 시간을 모두 중도절단으로 가정하는 것이다. 세 번째 상황은 생존 시간에는 독립이지만 공변량과는 독립이 아닌 중도절단 시간을 생성하였다(Schmid, Potapov 2012).

$i$  번째 대상의 생존 시간을  $T_i$ , 중도절단 시간을  $C_i$ 라고 정의하면 관찰된 생존 시간은  $X_i = \min(T_i, C_i)$ 로 정의할 수 있다. 중도절단 비율은 0, 0.3, 0.5, 0.7 총 네 가지 상황을 가정하였다. 실제 임상에서는 이보다 더 많은 대상이 중도절단 된 경우도 존재하기 때문에 중도절단 비율이 높은 0.7의 상황까지 가정하였다.

### 3.1.3 모의실험 설계

생성된 생존 시간과 중도절단 시간, 독립변수를 이용하여 Cox 비례위험모형을 이용하여 Harrell의  $C_{Harrell}$ , Heagerty의  $iAUC$ , Uno의  $C_\tau$ , Pencina의  $cNRI$ , Uno의  $NRI_{Uno}$ 의 성능을 1000번 반복 수행을 통해 비교한다. 모든 분석은 3.2.2 버전의 R 프로그램을 사용하여 분석하였으며, 다섯 가지 방법에 대해 survival, risksetROC, survC1, survIDINRI 패키지를 사용하였다. 실제 임상에서는 많은 자료가 얻어지는 경우가 적기 때문에 대상을 50, 100, 200의 상황으로 제한하였다. 중도절단 분포는 실제 임상에서 생길 수 있는 총 세 가지 상황을 반영하였으며, 회귀계수  $\beta$ 의 값을 통해 새로운 예측모형에 포함된 인자의 영향력을 조절하였다.  $C_{Harrell}$ ,  $iAUC$ ,  $C_\tau$ 는 기존 예측

모형과 새로운 예측모형에서 각각 구하여 차이를 구한 후, 얻어진 차이에 대한 95% 신뢰구간을 붓스트랩 방법을 이용하여 구한다.  $cNRI$ 와  $NRI_{uno}$ 는 두 예측모형을 이용하여 하나의 값을 구한 후, perturbation resampling 방법을 통해 각 값에 대한 95% 신뢰구간을 구한다. 각 방법들의 신뢰구간을 구하기 위해 붓스트랩과 perturbation resampling 방법은 각각 500번 수행하였다. 이렇게 구해진 다섯 가지 방법의 신뢰구간이 0 을 포함하는지 여부를 통해 성능을 비교 평가한다. 또한 추정된 다섯 가지 통계량의 Monte Carlo 표준편차(MCSD)를 근사적으로 추정하여 비교한다.

## 3.2 모의실험 결과

### 3.2.1 중도절단 형태가 랜덤한 경우

그림 1부터 3은 중도절단 형태가 랜덤한 경우를 나타낸다. 중도절단이 랜덤한 경우 Heagerty의  $iAUC$ 는 제 1종 오류가 매우 높게 나타났다. 두 예측모형에 유의한 차이가 없더라도 Heagerty의  $iAUC$ 는 유의한 결과를 나타낼 가능성이 매우 높을 것으로 예상된다. 하지만 Heagerty의  $iAUC$ 에 대한 제 1종 오류는 총 대상의 수가 많아질수록 5%에 가까워지는 것을 알 수 있었다. 모의실험에서는 총 대상의 수를 200으로 한정하였지만, 이보다 더 많은 대상으로 검정을 할 경우에는 Heagerty의  $iAUC$  역시 제 1종 오류가 5%에 가까울 것이라 판단된다. 또한, 총 대상의 수가 50인 경우를 제외하고는 중도절단의 비율이 높을수록 제 1종 오류가 안정되는 양상이 나타났다. 총 대상의 수가 200일 경우 Harrell의  $c$ -index는 모든 경우에서 제 1종 오류가 안정적으로 나타났으며, Uno의  $c$ -index는 오히려 제 1종 오류가 낮게 측정 되었다. 중도절단 비율이 0 일 경우에는 Harrell의  $c$ -index와 Uno의  $c$ -index가 검정력이 적절하게 나타났다. 하지만 중도절단 비율이 높아질수록 Uno의  $c$ -index는 검정력이 현저하게 낮아졌다. Pencina의  $NRI$ 와 Uno의  $NRI$ 는 모든 경우에서 Heagerty의  $iAUC$ 와는 반대로 제 1종 오류와 검정력이 너무 낮은 모습을 보였다. 즉, 두 예측모형의 예측력에 차이가 있어도

그 차이를 잘 잡아내지 못한다고 할 수 있다. Pencina의 NRI와 Uno의 NRI는 중도절단 비율이 높아질수록 검정력이 더 낮아지는 것을 확인할 수 있었다.

### 3.2.2 일정 시점 이후 모든 대상이 중도절단 된 경우

그림 4부터 6은 일정 시점 이후 모든 대상이 중도절단 된 경우를 나타낸다. 이 경우 역시 대상의 수가 많아질수록 Harrell의 c-index와 Uno의 c-index는 제 1종 오류가 5%에 가까워졌다. 특히 Uno의 c-index는 중도절단 비율이 높아져도 검정력 부분에서 매우 뛰어난 성능을 보였다. Harrell의 c-index는 중도절단이 랜덤한 경우와 비교했을 때, 변함없이 좋은 성능을 나타냈다. 하지만 중도절단 비율이 높아짐에 따라 Uno의 c-index가 조금 더 검정력이 높아지는 것을 확인할 수 있었다. Heagerty의  $i$ AUC는 대상의 수가 200인 경우에도 제 1종 오류가 5% 보다 높게 나타났다. Pencina와 Uno의 NRI는 중도절단이 랜덤한 경우에 비해 일정시점 이후 모든 대상이 중도절단 된 경우의 검정력이 뛰어났다. 특히, 대상의 수가 많아질수록 검정력이 확연하게 높아짐을 확인할 수 있었다. 하지만 제 1종 오류가 5% 보다 항상 작게 나타나기 때문에 두 예측모형의 예측력에 차이가 있을 경우 그 차이를 찾아내지 못하는 경우가 발생할 수 있을 것이다.

### 3.2.3 중도절단 시간이 공변량과 독립이 아닌 경우

그림 7부터 9는 중도절단 시간이 생존 시간  $T$ 와는 독립이면서 공변량과는 독립이 아닌 경우를 나타낸다. Harrell의 c-index는 대상의 수가 200일 경우 중도절단 비율과 상관없이 제 1종 오류가 5%에 가까웠다. 또한 회귀계수  $\beta$ 가 커질수록 검정력이 95% 보다 높게 나타났다. 하지만 중도절단 비율이 높을 경우에는 검정력이 다소 낮아짐을 확인할 수 있었다. Heagerty의  $i$ AUC는 여전히 제 1종 오류가 높게 나타나는 문제가 있었다. 두 예측모형의 예측력에 차이가 없더라도 Heagerty의  $i$ AUC는 차이가 있

다고 판단할 경우가 많이 발생할 것으로 예상된다. Pencina와 Uno의 NRI는 일정시점 이후 모든 대상이 중도절단된 경우와는 반대의 결과가 나타났다. 중도절단 비율이 높아짐에 따라 검정력이 현저히 낮아지는 것을 확인할 수 있었다. 또한, 다른 모든 경우와 마찬가지로 제 1종 오류는 모든 경우에서 5% 보다 작게 나타났다.





그림 1. 중도절단 형태가 랜덤하고 대상의 수가 50인 경우의 제 1종 오류와 검정력

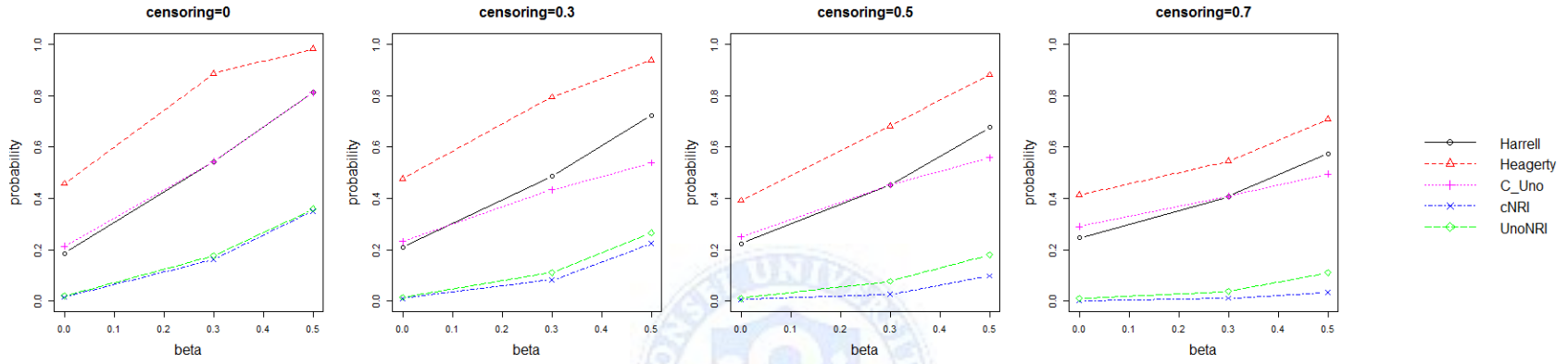


그림 2. 중도절단 형태가 랜덤하고 대상의 수가 100인 경우의 제 1종 오류와 검정력

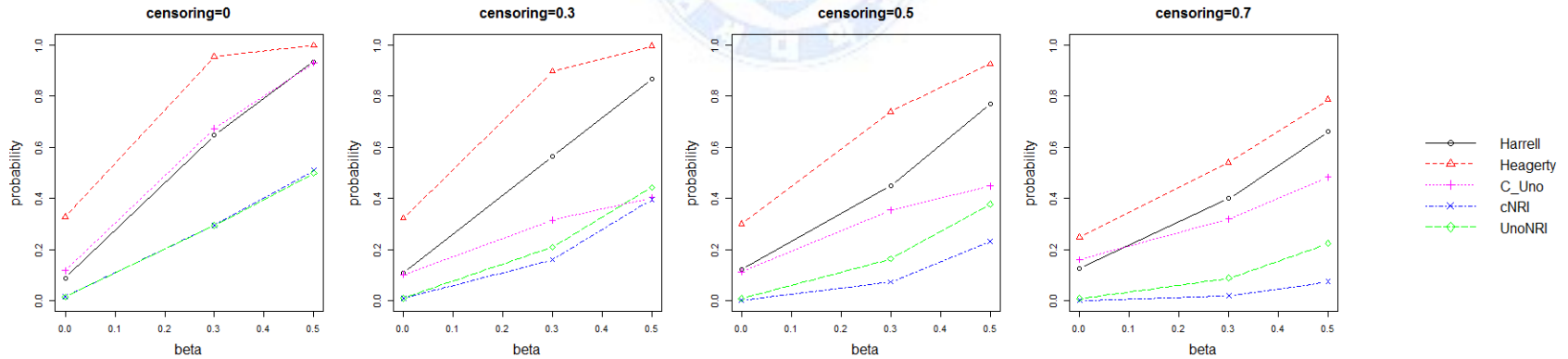


그림 3. 중도절단 형태가 랜덤하고 대상의 수가 200인 경우의 제 1종 오류와 검정력

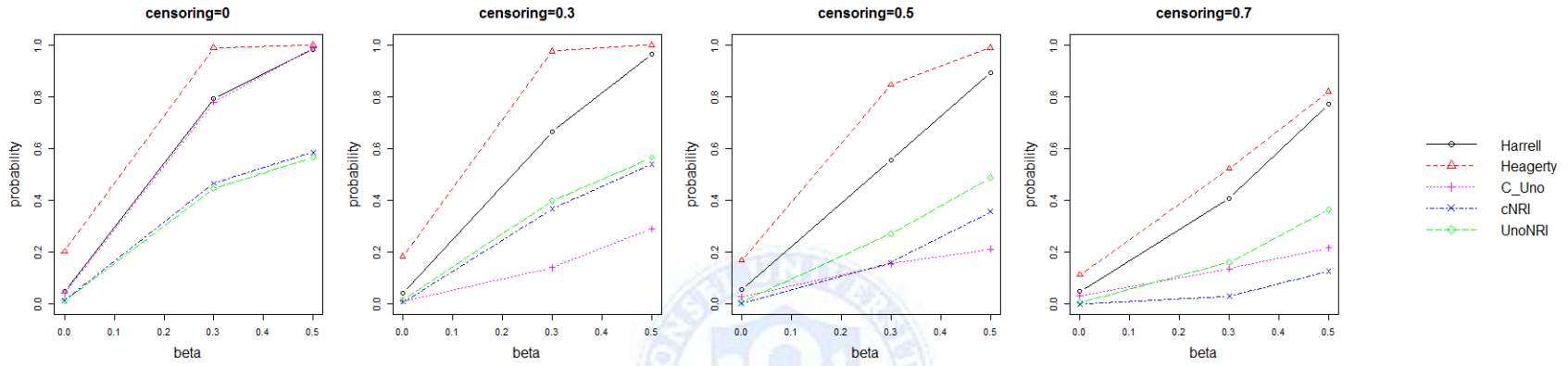


그림 4. 시점  $\tau$ 에서 모든 대상이 중도절단 되고 대상의 수가 50인 경우의 제 1종 오류와 검정력

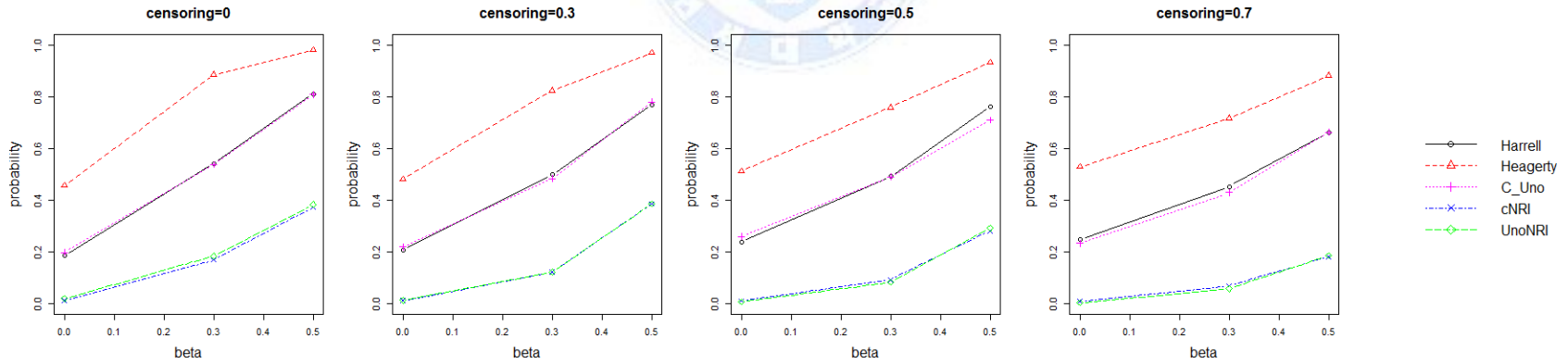


그림 5. 시점  $\tau$ 에서 모든 대상이 중도절단 되고 대상의 수가 100인 경우의 제 1종 오류와 검정력

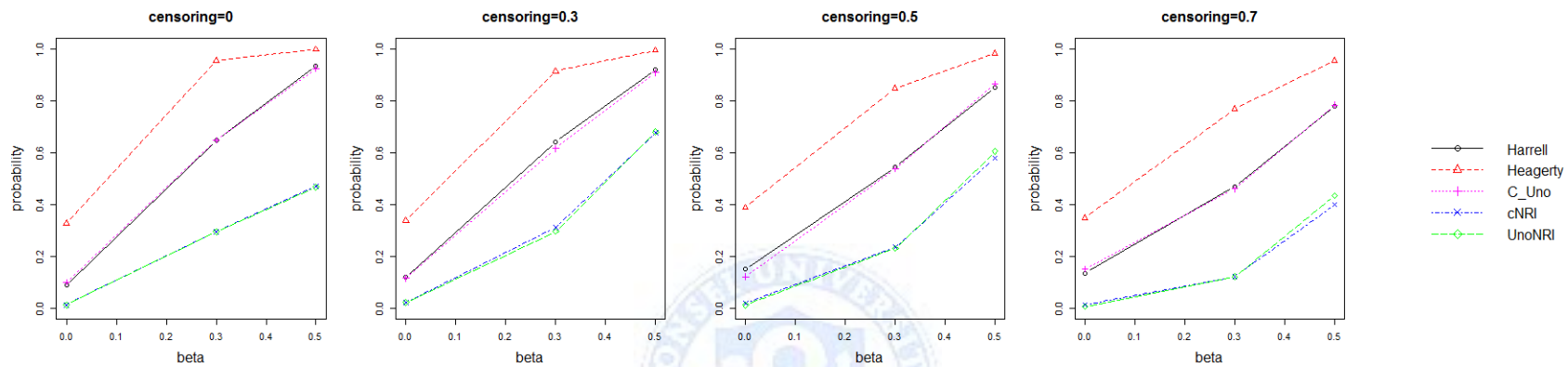


그림 6. 시점  $\tau$ 에서 모든 대상이 중도절단 되고 대상의 수가 200인 경우의 제 1종 오류와 검정력

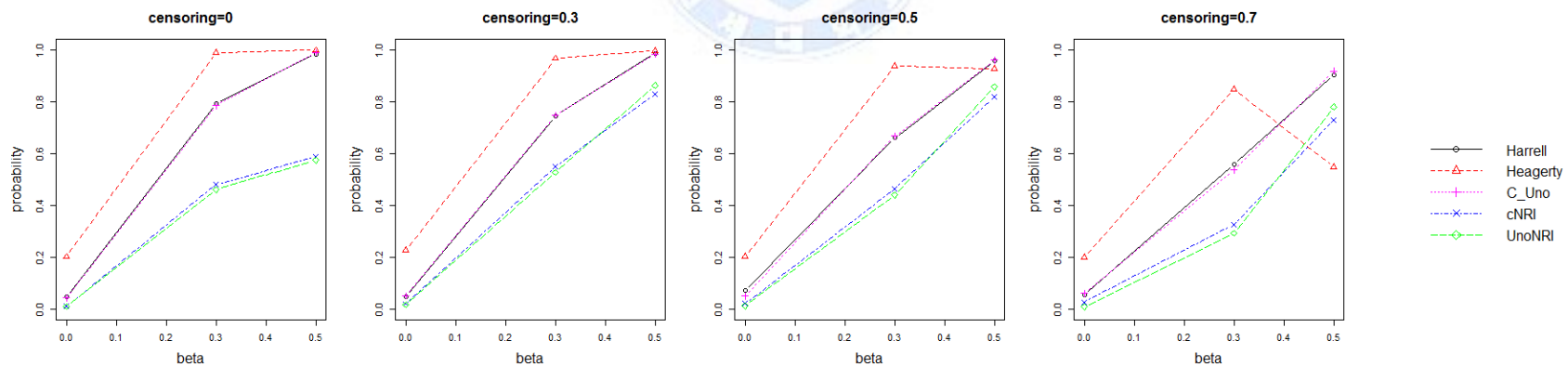


그림 7. 중도절단 시간이 생존 시간  $T$ 와는 독립이면서 공변량과는 독립이 아니고 대상의 수가 50인 경우의 제 1종 오류와 검정력

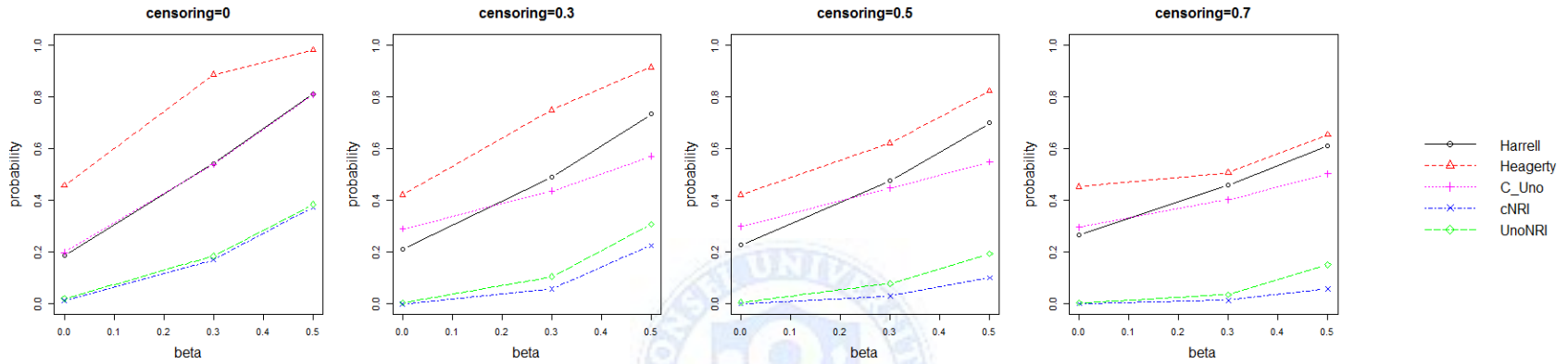


그림 8. 중도절단 시간이 생존 시간  $T$ 와는 독립이면서 공변량과는 독립이 아니고 대상의 수가 100인 경우의 제 1종 오류와 검정력

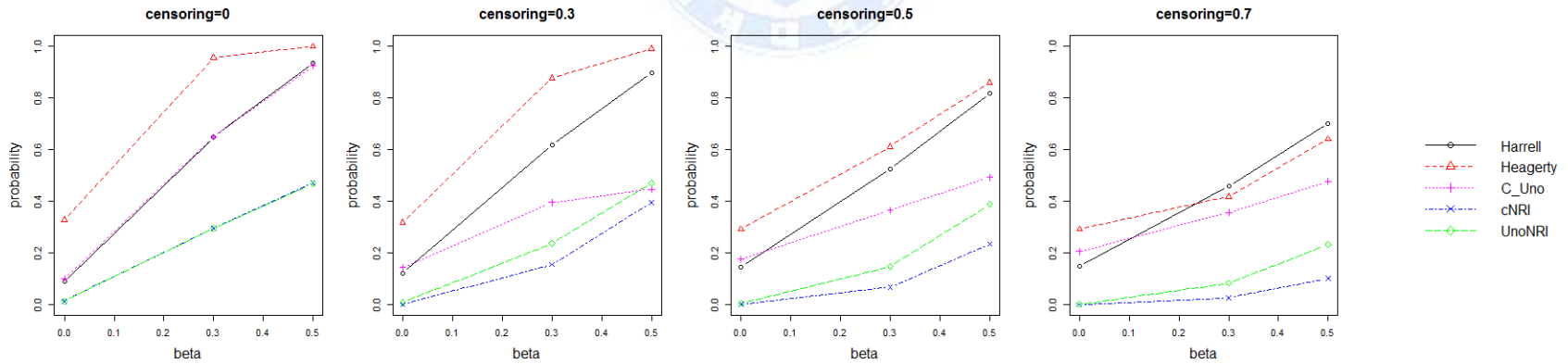
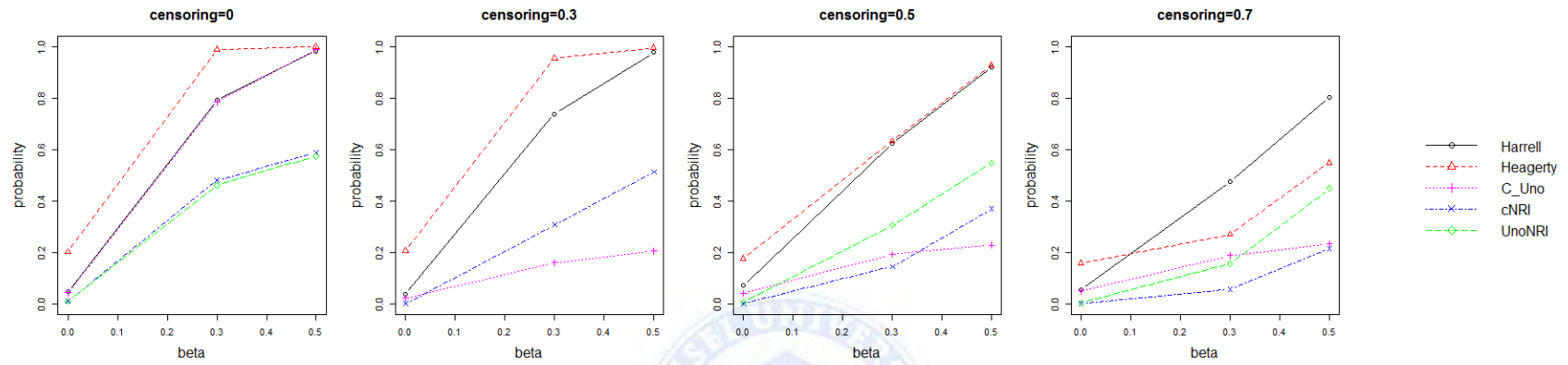


그림 9. 중도절단 시간이 생존 시간  $T$ 와는 독립이면서 공변량과는 독립이 아니고 대상의 수가 200인 경우의 제 1종 오류와 검정력



## 제 4 장 실제 자료 적용

본 연구는 생존 자료에서 기존 예측모형에 새로운 인자가 포함되었을 때 예측력이 향상 되었는지 판단하는 다섯 가지 통계량의 성능을 비교하는 것이 목적이다. 3장에서 모의실험을 통해 실제 임상에서 얻어질 수 있는 상황을 가정하여 비교를 해 보았으며, 4장에서는 실제 임상에서 사용된 자료를 이용하여 다섯 가지 통계량을 비교하고자 한다.

비교에 사용될 실제 자료는 Yonsei Stroke Registry에 2008년 11월부터 2010년 10월까지 등록된 환자들에 대해 후향적으로 연구한 Kim 외(2012)의 연구에 사용된 자료이다. 이 자료는 뇌졸중 발병으로부터 7일 이내에 응급실에 입원한 환자를 대상으로 874명의 자료를 수집하였다. Kim 외(2012) 연구의 목적은 적혈구분포도(Red blood cell Distribution Width; RDW)가 사망 또는 심혈관질환을 예측하는 인자로 적합한지 확인하는 것이다. 뇌졸중 발병환자의 모든 원인에 대한 사망을 예측하는 모형에 포함된 인자는 총 7 가지이며, 단계적 Cox 의 비례위험모형을 통해 선정 되었다. 연령, 성별, 최초 시점의 뇌졸중위험도평가점수(NIH Stroke Scale; NIHSS), 백혈구의 수, albumin, 잠재적 심인성 색전 요인이 있는 지 여부(the presence of high-risk potential cardiac sources of embolism), 악성 종양의 과거력이 예측모형에 포함되었다. 7 가지 인자가 포함된 예측모형에 RDW 인자가 포함되었을 때 뇌졸중 발병환자의 12개월 내 사망을 더 잘 예측할 수 있는 지 확인하였다.

두 예측모형의 예측력을 비교하기 위해 Heagerty의  $iAUC$ 를 이용하였으며, 기존 예측모형의  $iAUC$ 는 0.841(0.803-0.877), RDW가 포함된 예측모형의  $iAUC$ 는 0.858(0.822-0.892)가 도출 되었다. 두  $iAUC$ 의 차이에 대한 신뢰구간을 구하기 위해 붓스트랩을 1000번 수행하였고, 그 결과 차이에 대한 95% 신뢰구간은 0.004-0.035가 도출 되었다. 결론적으로 RDW가 포함된 예측모형의 예측력이 더 뛰어나다고 할 수 있었다.

본 연구에서는 Kim 외(2012)의 연구에 사용된 자료를 이용하여 2.2장에서 언급한

표 2. 실제 자료에 대한 다섯 가지 통계량의 결과

Method	Estimate	95% CI
$C_{Harrell}$	0.871-0.857=0.014	0.002-0.023
$iAUC$	0.858-0.841=0.017	0.003-0.034
$C_{uno}$	0.870-0.856=0.014	0.001-0.031
$cNRI$	0.304	0.036-0.433
$NRI_{uno}$	0.004	-0.007-0.020

다섯 가지 통계량을 비교한다. 분석에 사용될 자료는 91.5%가 중도절단 되었다. 최대 관찰 기간은 365일이며, 중도절단 대상 중 87.5%가 365일에서 중도절단 되었다. RDW가 포함되지 않은 예측모형과 포함된 예측모형에서 Harrell의 c-index, Heagerty의  $iAUC$ , Uno의 c-index를 각각 구하여 차이를 구하였다. 차이에 대한 95% 신뢰구간을 구하기 위해 붓스트랩을 1000번 수행하였다. 또한 두 예측모형의 재분류를 이용하여 Pencina의 cNRI, Uno의  $NRI_{uno}$ 를 구하였으며, perturbation resampling 방법을 이용하여 구해진 값의 95% 신뢰구간을 도출하였다. 그 결과는 표 2와 같다.

Harrell의 c-index, Heagerty의  $iAUC$ , Uno의 c-index, Pencina의 cNRI는 95% 신뢰구간이 0을 포함하지 않기 때문에 RDW가 포함된 새로운 예측모형의 예측력이 더 향상되었다는 결과가 도출 되었다. 반면 Uno의  $NRI_{uno}$ 는 95% 신뢰구간이 0을 포함하기 때문에 RDW가 포함된 예측모형이 기존의 예측모형에 비해 예측력이 향상되었다고 할 수 없었다.

## 제 5 장 결론

종속변수가 이분형 자료인 경우 독립이 아닌 두 예측모형에 대한 c-index, AUC의 비교는 두 예측모형에 대한 예측력의 작은 차이에 민감하게 반응하지 않는다고 알려져 있다(Cook 2007). 이러한 문제를 해결하기 위해 비교적 작은 차이에도 민감하게 반응할 수 있는 통계량인 NRI, IDI가 제안되었다(Pencina 외 2008). 두 예측모형에 대한 c-index, AUC의 비교와 NRI, IDI를 통한 예측력 향상을 평가하는 방법은 2.2에서와 같이 종속 변수가 생존 자료일 경우 그 특성을 반영할 수 있는 새로운 방법이 제안되어 왔다. 종속변수가 생존 자료일 때, 독립이 아닌 두 예측모형에 대한 예측력을 비교하기 위해 Harrell과 Uno는 c-index를 생존 자료의 특성에 맞게 수정하였으며, Heagerty는 모든 event 시점에서 AUC를 구한 후 적분하는 형태로 AUC를 변화시켰다. 또한, NRI 방법을 생존 자료의 특성에 맞게 수정하여 Pencina와 Uno가 제안하였다.

본 연구는 실제 임상에서 존재할 수 있는 여러 상황을 가정한 모의실험과 실제 자료를 이용하여, 생존 자료의 특성에 맞게 변화된 c-index, AUC, NRI 방법을 비교한 결과는 다음과 같다. 첫 째, 질병의 유무와 같은 이분형 자료인 경우와는 반대로, Pencina와 Uno의 NRI 방법이 Harrell과 Uno의 c-index와 Heagerty의  $i$ AUC 보다 두 예측모형의 작은 예측력 차이에 민감하게 반응하지 않았다. Harrell과 Uno의 c-index에 대한 제 1종 오류가 5%에 가까운 경우에, Pencina와 Uno의 NRI는 제 1종 오류가 현저히 작은 모습을 보였다. 생존 자료의 특성에 맞게 변화되면서 위와 같은 현상이 나타난 것으로 판단된다. 둘째, 다른 모든 경우에는 Harrell의 c-index가 가장 좋은 성능을 보인 반면, 일정 시점 이후에 모든 대상이 중도절단 된 경우에는 Uno의 c-index가 가장 좋은 성능을 보였다. Uno의 c-index는 중도절단 비율이 높아짐에도 불구하고 Uno의 c-index는 검정력이 90%를 상회하는 결과가 나타났다. 셋 째, Uno의 c-index의 성능에는 못 미치지만 일정 시점 이후 모든 대상이 중도절단 된 경우 Pencina와 Uno의 NRI 역시 다른 경우에 비해 좋은 성능을 보였다. 반대로 중도절단 형태가 랜덤한 경우와 생존 시간  $T$ 와는 독립이면서 공변량과는 독립이 아닌 경우



Pencina와 Uno의 NRI는 중도절단 비율이 높아질수록 검정력이 현저하게 낮아지는 결과가 나타났다.

실제 자료에서도 모의실험과 같은 결과가 나타나는지 보기 위해 사망 또는 심혈관 질환을 예측하는 인자로 적혈구분포도가 유의한 인자인지 판단하기 위해 수집한 뇌졸중 자료에 다섯 가지 방법을 적용해 본 결과 모의실험과 비슷한 결과를 얻을 수 있었다. 뇌졸중 자료는 91.5%가 중도절단되었으며, 최대 관찰 기간은 365일이다. 또한, 중도절단 대상 중 87.5%가 365일에서 중도절단 되었다. 일정 시점 이후 모든 대상이 중도절단 된 모의실험의 상황과 비슷하다. 적혈구분포도가 사망 또는 심혈관질환을 예측하는 유의한 인자인지 다섯 가지 방법으로 확인해 본 결과 Uno의 NRI를 제외한 모든 방법이 적혈구분포도가 포함된 예측모형의 예측력이 더 뛰어나다는 결과를 나타냈다. 실제로 Harrell의 c-index는 적혈구분포도 변수가 추가된 예측모형의 경우 0.871의 값을 나타내며, 적혈구분포도가 포함되지 않은 예측모형에 비해 평균적으로 0.014 정도 예측력이 향상되었다. 반면, Uno의 NRI는 95% 신뢰구간이 0을 포함하여 적혈구분포도가 포함된 예측모형과 포함되지 않은 예측모형에 예측력 차이가 없다는 결과가 나타났다. 모의실험과 실제 자료에 적용한 경우 모두 이분형 종속변수일 때와는 반대로 NRI가 예측력 차이에 민감하지 않은 결과를 나타냈다.

본 연구에서는 생존 자료에서 새로운 인자가 기존 예측모형에 추가되었을 때의 예측력 향상을 평가하는 방법으로 Harrell의 c-index, Heagerty의 *i*AUC, Uno의 c-index, Pencina의 NRI와 Uno의 NRI를 비교하였다. 본 연구에서 언급한 다섯 가지 방법 외에도 많은 방법이 연구되어 왔으며, 다른 방법과의 성능을 비교할 필요가 있다. 또한 더 다양한 상황을 가정하여 여러 방법들을 비교할 경우 본 연구와는 다른 결과가 나타날 수 있으므로, 향후 이 방면으로 많은 연구가 필요할 것으로 사료된다.

Bender, R., T. Augustin and M. Blettner. 2005. "Generating survival times to simulate Cox proportional hazards models". *Stat Med*, 24(11): 1713-23.

Cook, N. R. 2007. "Use and misuse of the receiver operating characteristic curve in risk prediction". *Circulation*, 115(7): 928-35.

Cox, D. R. 1972. "Regression models and life-tables". *Journal of the Royal Statistical Society. Series B (Methodological)*: 187-220.

DeLong, E. R., D. M. DeLong and D. L. Clarke-Pearson. 1988. "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach". *Biometrics*, 44(3): 837-45.

Efron, B. 1979. "Bootstrap methods: another look at the jackknife". *The Annals of Statistics*: 1-26.

Hajime Uno (2013). survC1: C-statistics for risk prediction models with censored survival data. R package version 1.0-2. <http://CRAN.R-project.org/package=survC1>

Hajime Uno and Tianxi Cai (2013). survIDINRI: IDI and NRI for comparing competing risk prediction models with censored survival data. R package version 1.1-1. <http://CRAN.R-project.org/package=survIDINRI>

Hanley, J. A. and B. J. McNeil. 1982. "The meaning and use of the area under a receiver operating characteristic (ROC) curve". *Radiology*, 143(1): 29-36.

Hanley, J. A. and B. J. McNeil. 1983. "A method of comparing the areas under receiver operating characteristic curves derived from the same cases". *Radiology*, 148(3): 839-43.

Harrell, F. E., Jr., R. M. Califf, D. B. Pryor, K. L. Lee and R. A. Rosati. 1982. "Evaluating the yield of medical tests". *JAMA*, 247(18): 2543-6.

Heagerty, P. J. and Y. Zheng. 2005. "Survival model predictive accuracy and ROC curves". *Biometrics*, 61(1): 92-105.

Kim J, Kim YD, Song TJ et al. Red blood cell distribution width is associated with poor clinical outcome in acute cerebral infarction. *Thromb Haemost.* 2012; 108: 349-356.

Patrick J. Heagerty and packaging by Paramita Saha-Chaudhuri (2012). risksetROC: Riskset ROC curve estimation from censored survival data. R package version 1.0.4. <http://CRAN.R-project.org/package=risksetROC>

Pencina, M. J., R. B. D'Agostino, Sr., R. B. D'Agostino, Jr. and R. S. Vasan. 2008. "Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond". *Stat Med*, 27(2): 157-72; discussion 207-12.

Pencina, M. J., R. B. D'Agostino, Sr. and E. W. Steyerberg. 2011. "Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers". *Stat Med*, 30(1): 11-21.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

Schmid, M. and S. Potapov. 2012. "A comparison of estimators to evaluate the discriminatory power of time-to-event models". *Stat Med*, 31(23): 2588-609.

Therneau T (2015). A Package for Survival Analysis in S. version 2.38, <http://CRAN.R-project.org/package=survival>. Terry M. Therneau and Patricia M. Grambsch (2000). Modeling Survival Data: Extending the Cox Model. Springer, New York. ISBN 0-387-98784-3.

Uno, H., T. Cai, M. J. Pencina, R. B. D'Agostino and L. J. Wei. 2011. "On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data". *Stat Med*, 30(10): 1105-17.

Uno, H., L. Tian, T. Cai, I. S. Kohane and L. J. Wei. 2013. "A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data". *Stat Med*, 32(14): 2430-42.



부록

A-1. 중도절단 형태가 랜덤하고, 중도절단 비율이 0인 경우의 각 방법에 대한 추정치 및 추정치의 표준오차, 제 1종 오류, 검정력

기존 인자와 새로 추가된 인자의 Cox 비례위험모형 회귀계수										
Sample size	Method	$\beta = (1, 0)$			$\beta = (1, 0.3)$			$\beta = (1, 0.5)$		
		Estimate	MCS D	type I error	Estimate	MCS D	Power	Estimate	MCS D	Power
50	$C_{Harrell}$	0.003	0.008	0.186	0.015	0.017	0.543	0.033	0.024	0.812
	$iAUC$	0.004	0.006	0.458	0.019	0.016	0.886	0.041	0.024	0.982
	$C_{Uno}$	0.004	0.008	0.214	0.015	0.017	0.543	0.033	0.024	0.812
	$cNRI$	-0.029	0.615	0.015	0.187	0.592	0.163	0.363	0.562	0.351
	$NRI_{Uno}$	0.002	0.015	0.021	0.012	0.032	0.177	0.026	0.047	0.359
100	$C_{Harrell}$	0.002	0.003	0.090	0.013	0.010	0.648	0.030	0.015	0.935
	$iAUC$	0.002	0.003	0.328	0.017	0.010	0.955	0.039	0.016	0.998
	$C_{Uno}$	0.002	0.003	0.118	0.014	0.011	0.674	0.032	0.016	0.930
	$cNRI$	-0.084	0.594	0.016	0.150	0.582	0.296	0.366	0.535	0.511
	$NRI_{Uno}$	0.001	0.010	0.015	0.010	0.029	0.295	0.024	0.044	0.500
200	$C_{Harrell}$	0.001	0.002	0.048	0.013	0.007	0.795	0.031	0.011	0.984
	$iAUC$	0.001	0.002	0.203	0.017	0.008	0.991	0.039	0.012	1.000
	$C_{Uno}$	0.001	0.002	0.041	0.013	0.007	0.781	0.031	0.011	0.991
	$cNRI$	-0.127	0.575	0.014	0.112	0.564	0.466	0.326	0.528	0.585
	$NRI_{Uno}$	0.001	0.007	0.013	0.009	0.027	0.445	0.023	0.042	0.566

A-2. 중도절단 형태가 랜덤하고, 중도절단 비율이 0.3인 경우의 각 방법에 대한 추정치 및 추정치의 표준오차, 제 1종 오류, 검정력

기존 인자와 새로 추가된 인자의 Cox 비례위험모형 회귀계수										
Sample size	Method	$\beta = (1, 0)$			$\beta = (1, 0.3)$			$\beta = (1, 0.5)$		
		Estimate	MCS D	type I error	Estimate	MCS D	Power	Estimate	MCS D	Power
50	$C_{Harrell}$	0.006	0.012	0.209	0.017	0.021	0.485	0.034	0.029	0.722
	$iAUC$	0.007	0.009	0.476	0.021	0.019	0.794	0.041	0.027	0.938
	$C_{Uno}$	0.010	0.140	0.232	0.025	0.188	0.434	0.038	0.263	0.539
	$cNRI$	0.022	0.618	0.011	0.261	0.574	0.082	0.404	0.543	0.224
	$NRI_{Uno}$	0.004	0.029	0.015	0.021	0.061	0.111	0.044	0.086	0.266
100	$C_{Harrell}$	0.003	0.005	0.108	0.015	0.013	0.565	0.033	0.020	0.869
	$iAUC$	0.003	0.004	0.324	0.017	0.012	0.897	0.038	0.019	0.996
	$C_{Uno}$	0.007	0.080	0.103	0.022	0.154	0.316	0.043	0.204	0.404
	$cNRI$	-0.004	0.598	0.009	0.234	0.576	0.161	0.435	0.527	0.396
	$NRI_{Uno}$	0.002	0.022	0.011	0.018	0.052	0.210	0.044	0.081	0.443
200	$C_{Harrell}$	0.001	0.002	0.043	0.013	0.009	0.666	0.031	0.013	0.966
	$iAUC$	0.002	0.002	0.183	0.016	0.008	0.977	0.037	0.012	1.000
	$C_{Uno}$	0.000	0.033	0.011	0.018	0.078	0.141	0.041	0.098	0.290
	$cNRI$	-0.060	0.590	0.008	0.253	0.559	0.367	0.459	0.500	0.541
	$NRI_{Uno}$	0.001	0.015	0.016	0.018	0.051	0.400	0.043	0.077	0.566

A-3. 중도절단 형태가 랜덤하고, 중도절단 비율이 0.5인 경우의 각 방법에 대한 추정치 및 추정치의 표준오차, 제 1종 오류, 검정력

		기존 인자와 새로 추가된 인자의 Cox 비례위험모형 회귀계수								
		$\beta = (1, 0)$			$\beta = (1, 0.3)$			$\beta = (1, 0.5)$		
Sample size	Method	Estimate	MCS D	type I error	Estimate	MCS D	Power	Estimate	MCS D	Power
50	$C_{Harrell}$	0.007	0.015	0.224	0.019	0.026	0.452	0.037	0.035	0.676
	$iAUC$	0.009	0.012	0.391	0.023	0.024	0.682	0.042	0.032	0.881
	$C_{Uno}$	0.014	0.177	0.251	0.031	0.221	0.452	0.042	0.282	0.559
	$cNRI$	0.102	0.608	0.006	0.255	0.583	0.026	0.410	0.544	0.098
	$NRI_{Uno}$	0.009	0.051	0.011	0.031	0.079	0.077	0.065	0.119	0.181
100	$C_{Harrell}$	0.004	0.008	0.123	0.014	0.015	0.451	0.031	0.023	0.770
	$iAUC$	0.004	0.005	0.300	0.016	0.014	0.739	0.035	0.021	0.926
	$C_{Uno}$	0.003	0.117	0.114	0.030	0.208	0.355	0.052	0.272	0.449
	$cNRI$	-0.003	0.600	0.002	0.279	0.572	0.074	0.462	0.514	0.233
	$NRI_{Uno}$	0.003	0.034	0.010	0.029	0.072	0.165	0.062	0.104	0.378
200	$C_{Harrell}$	0.002	0.003	0.056	0.014	0.011	0.555	0.032	0.016	0.893
	$iAUC$	0.002	0.003	0.168	0.015	0.010	0.847	0.035	0.014	0.991
	$C_{Uno}$	0.002	0.063	0.027	0.021	0.133	0.155	0.042	0.165	0.212
	$cNRI$	-0.014	0.585	0.002	0.268	0.565	0.159	0.432	0.528	0.357
	$NRI_{Uno}$	0.002	0.024	0.010	0.021	0.068	0.271	0.053	0.107	0.488

A-4. 중도절단 형태가 랜덤하고, 중도절단 비율이 0.7인 경우의 각 방법에 대한 추정치 및 추정치의 표준오차, 제 1종 오류, 검정력

기존 인자와 새로 추가된 인자의 Cox 비례위험모형 회귀계수										
Sample size	Method	$\beta = (1, 0)$			$\beta = (1, 0.3)$			$\beta = (1, 0.5)$		
		Estimate	MCS D	type I error	Estimate	MCS D	Power	Estimate	MCS D	Power
50	$C_{Harrell}$	0.014	0.026	0.246	0.025	0.037	0.407	0.043	0.048	0.575
	$iAUC$	0.016	0.024	0.414	0.027	0.038	0.545	0.045	0.048	0.708
	$C_{Uno}$	0.019	0.216	0.290	0.039	0.274	0.408	0.058	0.319	0.495
	$cNRI$	0.099	0.588	0.001	0.221	0.598	0.012	0.377	0.560	0.033
	$NRI_{Uno}$	0.010	0.062	0.010	0.038	0.098	0.038	0.081	0.130	0.111
100	$C_{Harrell}$	0.006	0.012	0.128	0.018	0.022	0.401	0.036	0.030	0.661
	$iAUC$	0.008	0.012	0.248	0.019	0.021	0.541	0.037	0.030	0.787
	$C_{Uno}$	0.013	0.152	0.161	0.029	0.209	0.320	0.051	0.277	0.484
	$cNRI$	0.050	0.591	0.002	0.281	0.557	0.020	0.436	0.516	0.075
	$NRI_{Uno}$	0.006	0.043	0.009	0.033	0.087	0.089	0.074	0.118	0.225
200	$C_{Harrell}$	0.003	0.005	0.048	0.016	0.014	0.409	0.034	0.021	0.772
	$iAUC$	0.004	0.005	0.113	0.016	0.014	0.523	0.034	0.020	0.822
	$C_{Uno}$	0.000	0.087	0.033	0.023	0.168	0.137	0.050	0.213	0.217
	$cNRI$	0.063	0.585	0.000	0.279	0.551	0.031	0.449	0.502	0.126
	$NRI_{Uno}$	0.004	0.030	0.007	0.029	0.074	0.161	0.070	0.106	0.365



B-1. 시점  $\tau$ 에서 모든 대상이 중도절단 되고, 중도절단 비율이 0인 경우의 각 방법에 대한 추정치 및 추정치의 표준오차, 제 1종 오류, 검정력

기존 인자와 새로 추가된 인자의 Cox 비례위험모형 회귀계수										
Sample size	Method	$\beta = (1, 0)$			$\beta = (1, 0.3)$			$\beta = (1, 0.5)$		
		Estimate	MCS D	type I error	Estimate	MCS D	Power	Estimate	MCS D	Power
50	$C_{Harrell}$	0.003	0.008	0.186	0.015	0.017	0.543	0.033	0.024	0.812
	$iAUC$	0.004	0.006	0.458	0.019	0.016	0.886	0.041	0.024	0.982
	$C_{Uno}$	0.003	0.008	0.199	0.015	0.017	0.538	0.033	0.024	0.809
	$cNRI$	0.008	0.600	0.013	0.197	0.593	0.170	0.386	0.554	0.373
	$NRI_{Uno}$	0.002	0.015	0.020	0.013	0.033	0.185	0.028	0.047	0.384
100	$C_{Harrell}$	0.002	0.003	0.090	0.013	0.010	0.648	0.030	0.015	0.935
	$iAUC$	0.002	0.003	0.328	0.017	0.010	0.955	0.039	0.016	0.998
	$C_{Uno}$	0.002	0.004	0.101	0.013	0.010	0.649	0.031	0.015	0.925
	$cNRI$	-0.049	0.589	0.014	0.186	0.580	0.296	0.359	0.546	0.471
	$NRI_{Uno}$	0.001	0.010	0.014	0.010	0.028	0.296	0.022	0.043	0.468
200	$C_{Harrell}$	0.001	0.002	0.048	0.013	0.007	0.795	0.031	0.011	0.984
	$iAUC$	0.001	0.002	0.203	0.017	0.008	0.991	0.039	0.012	0.999
	$C_{Uno}$	0.001	0.002	0.045	0.013	0.008	0.786	0.031	0.011	0.990
	$cNRI$	-0.103	0.575	0.013	0.150	0.554	0.482	0.337	0.523	0.589
	$NRI_{Uno}$	0.001	0.006	0.011	0.009	0.025	0.462	0.021	0.039	0.575

B-2. 시점  $\tau$ 에서 모든 대상이 중도절단 되고, 중도절단 비율이 0.3인 경우의 각 방법에 대한 추정치 및 추정치의 표준오차, 제 1종 오류, 검정력

기존 인자와 새로 추가된 인자의 Cox 비례위험모형 회귀계수										
Sample size	Method	$\beta = (1, 0)$			$\beta = (1, 0.3)$			$\beta = (1, 0.5)$		
		Estimate	MCS D	type I error	Estimate	MCS D	Power	Estimate	MCS D	Power
50	$C_{Harrell}$	0.005	0.010	0.207	0.016	0.018	0.499	0.033	0.026	0.771
	$iAUC$	0.006	0.008	0.482	0.019	0.017	0.824	0.038	0.024	0.971
	$C_{Uno}$	0.005	0.009	0.220	0.016	0.018	0.484	0.033	0.026	0.781
	$cNRI$	0.075	0.504	0.012	0.311	0.458	0.123	0.496	0.353	0.387
	$NRI_{Uno}$	0.003	0.036	0.013	0.044	0.070	0.123	0.104	0.102	0.387
100	$C_{Harrell}$	0.002	0.004	0.122	0.014	0.012	0.642	0.032	0.017	0.920
	$iAUC$	0.003	0.003	0.339	0.016	0.011	0.915	0.036	0.017	0.995
	$C_{Uno}$	0.003	0.004	0.117	0.014	0.012	0.617	0.032	0.018	0.909
	$cNRI$	0.045	0.502	0.023	0.342	0.433	0.314	0.526	0.318	0.677
	$NRI_{Uno}$	0.001	0.025	0.024	0.039	0.059	0.297	0.103	0.092	0.681
200	$C_{Harrell}$	0.001	0.002	0.049	0.013	0.008	0.746	0.030	0.012	0.987
	$iAUC$	0.002	0.002	0.228	0.014	0.008	0.967	0.037	0.013	0.997
	$C_{Uno}$	0.001	0.002	0.053	0.013	0.008	0.749	0.030	0.012	0.983
	$cNRI$	0.063	0.502	0.022	0.397	0.397	0.550	0.566	0.262	0.830
	$NRI_{Uno}$	0.001	0.013	0.018	0.036	0.045	0.528	0.093	0.073	0.864

B-3. 시점  $\tau$ 에서 모든 대상이 중도절단 되고, 중도절단 비율이 0.5인 경우의 각 방법에 대한 추정치 및 추정치의 표준오차, 제 1종 오류, 검정력

기존 인자와 새로 추가된 인자의 Cox 비례위험모형 회귀계수										
Sample size	Method	$\beta = (1, 0)$			$\beta = (1, 0.3)$			$\beta = (1, 0.5)$		
		Estimate	MCS D	type I error	Estimate	MCS D	Power	Estimate	MCS D	Power
50	$C_{Harrell}$	0.006	0.011	0.239	0.017	0.020	0.493	0.034	0.028	0.764
	$iAUC$	0.007	0.010	0.513	0.019	0.019	0.760	0.037	0.027	0.934
	$C_{Uno}$	0.007	0.012	0.261	0.017	0.021	0.491	0.034	0.029	0.711
	$cNRI$	0.122	0.503	0.011	0.287	0.475	0.093	0.484	0.360	0.282
	$NRI_{Uno}$	0.008	0.028	0.008	0.037	0.057	0.084	0.088	0.085	0.293
100	$C_{Harrell}$	0.003	0.005	0.151	0.014	0.013	0.544	0.030	0.018	0.852
	$iAUC$	0.004	0.004	0.389	0.015	0.012	0.849	0.034	0.018	0.984
	$C_{Uno}$	0.003	0.005	0.122	0.013	0.013	0.538	0.031	0.019	0.866
	$cNRI$	0.061	0.504	0.020	0.372	0.409	0.236	0.523	0.304	0.580
	$NRI_{Uno}$	0.002	0.017	0.013	0.036	0.044	0.232	0.084	0.068	0.606
200	$C_{Harrell}$	0.002	0.003	0.073	0.013	0.009	0.661	0.030	0.013	0.958
	$iAUC$	0.002	0.002	0.204	0.014	0.008	0.938	0.032	0.016	0.929
	$C_{Uno}$	0.002	0.002	0.052	0.013	0.009	0.668	0.030	0.013	0.961
	$cNRI$	0.064	0.500	0.023	0.396	0.389	0.465	0.553	0.255	0.819
	$NRI_{Uno}$	0.001	0.011	0.014	0.031	0.035	0.441	0.077	0.053	0.858

B-4. 시점  $\tau$ 에서 모든 대상이 중도절단 되고, 중도절단 비율이 0.7인 경우의 각 방법에 대한 추정치 및 추정치의 표준오차, 제 1종 오류, 검정력

기존 인자와 새로 추가된 인자의 Cox 비례위험모형 회귀계수										
Sample size	Method	$\beta = (1, 0)$			$\beta = (1, 0.3)$			$\beta = (1, 0.5)$		
		Estimate	MCS D	type I error	Estimate	MCS D	Power	Estimate	MCS D	Power
50	$C_{Harrell}$	0.010	0.017	0.248	0.019	0.026	0.454	0.035	0.035	0.664
	$iAUC$	0.011	0.013	0.530	0.021	0.023	0.717	0.039	0.032	0.883
	$C_{Uno}$	0.009	0.017	0.234	0.018	0.022	0.430	0.035	0.031	0.664
	$cNRI$	0.131	0.506	0.009	0.299	0.467	0.070	0.469	0.369	0.182
	$NRI_{Uno}$	0.010	0.028	0.004	0.033	0.050	0.058	0.071	0.074	0.187
100	$C_{Harrell}$	0.005	0.007	0.137	0.015	0.015	0.470	0.031	0.021	0.781
	$iAUC$	0.005	0.006	0.350	0.016	0.014	0.770	0.034	0.020	0.955
	$C_{Uno}$	0.005	0.008	0.153	0.015	0.015	0.461	0.031	0.022	0.784
	$cNRI$	0.069	0.505	0.015	0.360	0.414	0.123	0.519	0.291	0.401
	$NRI_{Uno}$	0.004	0.014	0.008	0.026	0.031	0.122	0.063	0.051	0.435
200	$C_{Harrell}$	0.003	0.003	0.056	0.013	0.010	0.560	0.029	0.015	0.903
	$iAUC$	0.003	0.003	0.201	0.014	0.010	0.840	0.027	0.024	0.550
	$C_{Uno}$	0.003	0.004	0.062	0.013	0.010	0.537	0.030	0.015	0.918
	$cNRI$	0.039	0.502	0.028	0.373	0.399	0.327	0.541	0.257	0.729
	$NRI_{Uno}$	0.002	0.009	0.010	0.024	0.025	0.294	0.060	0.040	0.781

C-1. 중도절단 시간이 생존 시간  $T$ 와는 독립이면서 공변량과는 독립이 아닐 때, 중도절단 비율이 0인 경우의 각 방법에 대한 추정치 및 추정치의 표준오차, 제 1종 오류, 검정력

		기존 인자와 새로 추가된 인자의 Cox 비례위험모형 회귀계수								
		$\beta = (1, 0)$			$\beta = (1, 0.3)$			$\beta = (1, 0.5)$		
Sample size	Method	Estimate	MCS D	type I error	Estimate	MCS D	Power	Estimate	MCS D	Power
50	$C_{Harrell}$	0.003	0.008	0.186	0.015	0.017	0.543	0.033	0.024	0.812
	$iAUC$	0.004	0.006	0.458	0.019	0.016	0.886	0.041	0.024	0.982
	$C_{Uno}$	0.003	0.008	0.199	0.015	0.017	0.538	0.033	0.024	0.809
	$cNRI$	0.008	0.600	0.013	0.197	0.593	0.170	0.386	0.554	0.373
	$NRI_{Uno}$	0.002	0.015	0.020	0.013	0.033	0.185	0.028	0.047	0.384
100	$C_{Harrell}$	0.002	0.003	0.090	0.013	0.010	0.648	0.030	0.015	0.935
	$iAUC$	0.002	0.003	0.328	0.017	0.010	0.955	0.039	0.016	0.998
	$C_{Uno}$	0.002	0.004	0.101	0.013	0.010	0.649	0.031	0.015	0.925
	$cNRI$	-0.049	0.589	0.014	0.186	0.580	0.296	0.359	0.546	0.471
	$NRI_{Uno}$	0.001	0.010	0.014	0.010	0.028	0.296	0.022	0.043	0.468
200	$C_{Harrell}$	0.001	0.002	0.048	0.013	0.007	0.795	0.031	0.011	0.984
	$iAUC$	0.001	0.002	0.203	0.017	0.008	0.991	0.039	0.012	1.000
	$C_{Uno}$	0.001	0.002	0.045	0.013	0.008	0.786	0.031	0.011	0.990
	$cNRI$	-0.103	0.575	0.013	0.150	0.554	0.482	0.337	0.523	0.589
	$NRI_{Uno}$	0.001	0.006	0.011	0.009	0.025	0.462	0.021	0.039	0.575

C-2. 중도절단 시간이 생존 시간  $T$ 와는 독립이면서 공변량과는 독립이 아닐 때, 중도절단 비율이 0.3인 경우의 각 방법에 대한 추정치 및 추정치의 표준오차, 제 1종 오류, 검정력

		기존 인자와 새로 추가된 인자의 Cox 비례위험모형 회귀계수								
		$\beta = (1, 0)$			$\beta = (1, 0.3)$			$\beta = (1, 0.5)$		
Sample size	Method	Estimate	MCS D	type I error	Estimate	MCS D	Power	Estimate	MCS D	Power
50	$C_{Harrell}$	0.006	0.011	0.212	0.018	0.021	0.489	0.035	0.029	0.734
	$iAUC$	0.007	0.010	0.421	0.021	0.021	0.749	0.041	0.029	0.915
	$C_{Uno}$	0.012	0.212	0.290	0.035	0.295	0.436	0.034	0.336	0.571
	$cNRI$	0.059	0.583	0.000	0.329	0.552	0.590	0.516	0.486	0.225
	$NRI_{Uno}$	0.006	0.046	0.004	0.040	0.085	0.105	0.082	0.115	0.308
100	$C_{Harrell}$	0.003	0.005	0.123	0.016	0.013	0.618	0.035	0.019	0.897
	$iAUC$	0.004	0.005	0.319	0.018	0.013	0.876	0.039	0.019	0.990
	$C_{Uno}$	0.015	0.154	0.144	0.038	0.279	0.396	0.048	0.328	0.447
	$cNRI$	0.018	0.573	0.001	0.365	0.537	0.156	0.571	0.451	0.396
	$NRI_{Uno}$	0.002	0.031	0.009	0.033	0.073	0.238	0.076	0.106	0.471
200	$C_{Harrell}$	0.002	0.002	0.040	0.014	0.009	0.740	0.033	0.013	0.978
	$iAUC$	0.002	0.002	0.207	0.016	0.009	0.955	0.037	0.013	0.997
	$C_{Uno}$	-0.001	0.100	0.024	0.015	0.211	0.162	0.029	0.247	0.207
	$cNRI$	-0.013	0.562	0.002	0.394	0.502	0.309	0.582	0.435	0.514
	$NRI_{Uno}$	0.001	0.022	0.010	0.034	0.073	0.408	0.081	0.111	0.577

C-3. 중도절단 시간이 생존 시간  $T$ 와는 독립이면서 공변량과는 독립이 아닐 때, 중도절단 비율이 0.5인 경우의 각 방법에 대한 추정치 및 추정치의 표준오차, 제 1종 오류, 검정력

기존 인자와 새로 추가된 인자의 Cox 비례위험모형 회귀계수										
Sample size	Method	$\beta = (1, 0)$			$\beta = (1, 0.3)$			$\beta = (1, 0.5)$		
		Estimate	MCS D	type I error	Estimate	MCS D	Power	Estimate	MCS D	Power
50	$C_{Harrell}$	0.008	0.015	0.227	0.021	0.026	0.476	0.040	0.036	0.699
	$iAUC$	0.011	0.015	0.421	0.023	0.027	0.622	0.042	0.037	0.823
	$C_{Uno}$	0.024	0.234	0.300	0.048	0.326	0.448	0.066	0.348	0.549
	$cNRI$	0.031	0.575	0.001	0.352	0.542	0.031	0.538	0.451	0.101
	$NRI_{Uno}$	0.007	0.051	0.006	0.047	0.090	0.078	0.094	0.125	0.194
100	$C_{Harrell}$	0.004	0.007	0.146	0.017	0.016	0.526	0.035	0.023	0.816
	$iAUC$	0.006	0.008	0.293	0.017	0.017	0.612	0.034	0.023	0.860
	$C_{Uno}$	0.022	0.188	0.177	0.038	0.290	0.366	0.044	0.336	0.493
	$cNRI$	0.016	0.560	0.002	0.373	0.521	0.068	0.560	0.446	0.235
	$NRI_{Uno}$	0.002	0.031	0.006	0.040	0.078	0.148	0.096	0.116	0.389
200	$C_{Harrell}$	0.003	0.003	0.073	0.015	0.011	0.626	0.034	0.016	0.921
	$iAUC$	0.003	0.004	0.176	0.014	0.011	0.634	0.032	0.016	0.929
	$C_{Uno}$	0.008	0.120	0.042	0.024	0.243	0.194	0.036	0.286	0.230
	$cNRI$	0.006	0.556	0.003	0.427	0.493	0.146	0.609	0.411	0.371
	$NRI_{Uno}$	0.001	0.024	0.008	0.039	0.073	0.308	0.091	0.108	0.549

C-4. 중도절단 시간이 생존 시간  $T$ 와는 독립이면서 공변량과는 독립이 아닐 때, 중도절단 비율이 0.7인 경우의 각 방법에 대한 추정치 및 추정치의 표준오차, 제 1종 오류, 검정력

		기존 인자와 새로 추가된 인자의 Cox 비례위험모형 회귀계수								
		$\beta = (1, 0)$			$\beta = (1, 0.3)$			$\beta = (1, 0.5)$		
Sample size	Method	Estimate	MCS D	type I error	Estimate	MCS D	Power	Estimate	MCSE	Power
50	$C_{Harrell}$	0.016	0.028	0.267	0.029	0.040	0.459	0.047	0.051	0.611
	$iAUC$	0.021	0.032	0.453	0.029	0.044	0.507	0.045	0.051	0.654
	$C_{Uno}$	0.018	0.243	0.296	0.050	0.309	0.403	0.081	0.337	0.503
	$cNRI$	0.103	0.560	0.001	0.321	0.552	0.014	0.526	0.451	0.058
	$NRI_{Uno}$	0.011	0.058	0.004	0.050	0.095	0.035	0.115	0.139	0.152
100	$C_{Harrell}$	0.007	0.011	0.148	0.020	0.021	0.458	0.039	0.030	0.700
	$iAUC$	0.011	0.017	0.293	0.018	0.026	0.419	0.033	0.032	0.641
	$C_{Uno}$	0.021	0.191	0.207	0.033	0.276	0.357	0.065	0.320	0.477
	$cNRI$	0.070	0.546	0.001	0.400	0.504	0.026	0.565	0.419	0.102
	$NRI_{Uno}$	0.005	0.035	0.003	0.047	0.068	0.084	0.103	0.101	0.233
200	$C_{Harrell}$	0.004	0.005	0.057	0.017	0.014	0.477	0.036	0.020	0.805
	$iAUC$	0.006	0.010	0.159	0.013	0.017	0.272	0.027	0.024	0.550
	$C_{Uno}$	0.006	0.129	0.051	0.031	0.230	0.189	0.047	0.269	0.235
	$cNRI$	0.008	0.545	0.002	0.411	0.491	0.057	0.580	0.406	0.216
	$NRI_{Uno}$	0.001	0.023	0.004	0.040	0.066	0.157	0.103	0.098	0.450



## ABSTRACT

### Comparison of measures evaluating performance for a new factor in survival data

Chu, Soon Gyu

Dept. of Biostatistics and Computing

The Graduate School

Yonsei University

There are a variety of statistical methods to predict the presence or absence of disease in medical research. When a new factor is added to the existing prediction model, the improvement of the predictive power can be evaluated using  $c$ -index or net reclassification index (NRI). For binary outcome data, a number of studies show that NRI is more sensitive to detect an improvement in predictive ability of the added factor. For survival outcome data, however, such an evaluation has not been done thoroughly. In this study, we compare several methods estimating the improvement of the added predictive power for survival outcome data through a simulation study under various scenarios. We consider  $c$ -index (Harrell, 1982), iAUC (integrated area under the ROC curve) (Heagerty, 2005),  $c$ -index (Uno, 2011), NRI (Pencina, 2011), and NRI (Uno, 2013). The simulation study results demonstrate that there is an inflation of the type I error rate for  $c$ -index and iAUC, while NRIs are very conservative. We also compare the results of the five methods applied to a real data example.

---

Key words :  $c$ -index, iAUC, NRI, survival analysis, predicted model