



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](#)

유방암 수술 경험자에서 암 재발
위험요인 및 재발 예측모델 연구

연세대학교 보건대학원
보건통계학과 보건통계전공
이 해 림

유방암 수술 경험자에서 암 재발 위험요인 및 재발 예측모델 연구

지도 남 정 모 교수

이 논문을 보건학 석사학위 논문으로 제출함.

2024년 6월 일

연세대학교 보건대학원
보건통계학과 보건통계전공
이 해 립

이해림의 보건학 석사학위 논문을 인준함.

심사위원 남 정 모 인

심사위원 박 소 희 인

심사위원 박 형 석 인

연세대학교 보건대학원

2024년 6월 일

감사의 말씀

“We can make our plans, but the Lord determines our steps.”

(Proverbs 16:9)

제가 꿈꾸고 생각했던 것 이상으로 항상 인도해주시는 하나님 감사합니다.

수술실 간호사로 일이 익숙해지고 새로운 도전을 찾고 있던 무렵 보건 통계에 관심을 가지게 되었고 진학한 대학원을 어느덧 졸업을 하게 되었습니다. 학부에서 많이 다루지 않았던 통계를 전공하려니 시작은 조금 막막했으나 많은 분들의 도움과 지지를 받아 여기까지 올 수 있었습니다.

때로는 따뜻하게 때로는 냉철하게 논문을 지도해 주신 든든한 남정모 교수님, 따뜻하고 명쾌한 강의력으로 많은 가르침을 주신 박소희 교수님께 감사드립니다. 바쁘신 와중에 논문심사를 맡아주시고 데이터 제공에도 도움을 주신 박형석 교수님께 감사드립니다.

새로운 도전을 응원하고 늘 지지해주시는 사랑하는 아빠, 늘 기도해주시고 신앙적으로 올바른 길을 알려주시는 사랑하는 엄마 감사합니다. 논문의 어려움을 누구보다 잘 알고 지지해준 해진이언니, 자상하게 응원해주던 상웅이오빠, 우울할 때 함께 놀아준 해연이언니, 입학시기가 같아 함께 공부한 상찬이까지 사랑하는 독수리 오남매 감사합니다. 바라만 보아도 기분좋아지는 사랑하는 조카들 재희, 재윤, 창주, 윤주, 주영아 이모(고모) 석사 졸업했다!

힘들 때 맛있는 거 사주고 응원해준 사랑하는 병관이 나의 스트레스를 다 받아주느라 고생했어. 고맙고 미안하고 사랑합니다.

대학원 수업을 위해 오버타임 근무와 이브닝 근무를 바꿔야 할 때 흔쾌히 바꿔주신 신촌세브란스 암병원 로봇방 선생님들께도 감사의 말을 전합니다. 일과 학업에 지칠 때마다 든든한 지지체계가 되어준 소중한 동기들 다혜 언니, 유경 언니, 진영이 고맙습니다.

대학원 진학 준비부터 함께 기도해주고 응원해주던 준형셀, 논문 학기 동안 함께 마음 졸여주던 선영셀, 한결셀 너무 고맙습니다. 여러분들의 기도 덕분에입니다! 벌써 두 아이의 엄마로 너무나 멋지게 살아가는 주향이, 새로운 곳에서도 열정적으로 일하고 있는 온유, 항상 잘 들어주는 공감 요정 서연 언니, 멋지다는 칭찬으로 늘 응원해주는 민아까지 사랑하는 칭둥이들 고맙습니다.

2024년 8월, 대학원 생활에 마침표를 찍고 앞으로 1년은 마다가스카르 의료 선교라는 새로운 발걸음을 떼려 갑니다. 대학원 생활을 통해 얻은 지식과 추억을 잘 간직하고 간호와 통계를 이용하여 그 곳에 여러 도움을 드릴 수 있는 제가 되겠습니다. 함께해주신 모든 분들께 다시 한번 감사의 말을 전합니다.

2024년 6월

이해림 올림

차 례

국문요약

I. 서론

- 1. 연구의 배경 및 필요성 1
- 2. 연구의 목적 3

II. 이론적 배경

- 1. 유방암 4
- 2. 기계학습 6

III. 연구 방법

- 1. 연구 모형 9
- 2. 연구 대상 및 자료원 10
- 3. 변수의 선정 및 정의 10
- 4. 분석 방법 14
- 5. 윤리적 고려사항 15

IV. 연구 결과

- 1. 일반적 특성16
 - 1.1. 유방암 재발 여부에 대한 일반적 특성16
 - 1.2. 유방암 분자 아형에 따른 일반적 특성20

2. 유방암 수술 후 암 재발률 및 재발 비례위험도	24
2.1. 유방암 수술 후 무 재발 생존에 영향을 미치는 요인	24
2.2. 유방암 수술 후 암 재발 상대위험도	31
3. 유방암 분자 아형에 따른 암 재발 위험 하위그룹 분석	35
4. Cox Proportional Hazards Model, Random Survival forest, XGBSE를 이용한 재발예측 모델 비교	39
 V. 고찰	
1. 연구 결과에 대한 고찰	42
2. 연구의 강점과 제한점	45
 VI. 결론	
 참고 문헌	
 부록	
 영문 요약	

표 차 례

Table 1. Baseline characteristics by Breast cancer recurrence	18
Table 2. Baseline characteristics by Breast cancer subtype	22
Table 3. Recurrence-Free Survival Rates of 3-Year, 5-Year, 10-Year	26
Table 4. Hazard ratio for Recurrence of Breast cancer	33
Table 5. C-index of RSF Model, XGBSE Model	40
Table 6. Best Parameter of RSF Model, XGBSE Model	40

그 림 차 례

Figure 1. 유방암 증가 현황	1
Figure 2. Framework of study	9
Figure 3. Kaplan-Meier Curve & Log rank Test	28
Figure 4. HR+/HER2-대비 HR+/HER2+ HR forest plot	36
Figure 5. HR+/HER2-대비 HR-/HER2+ HR forest plot	37
Figure 6. HR+/HER2-대비 HR-/HER2- HR forest plot	38
Figure 7. Feature Importance plot of RSF, XGBSE Model	41

부 록 차 례

부록 1. Kaplan-Meier Curve & Log rank Test	52
부록 2. Best Parameter of RSF Model, XGBSE Model	55

국 문 요 약

유방암 수술 경험자에서 암 재발 위험요인 및 재발예측 모델 연구

조기 발견 및 개선된 의료 기술로 유방암 생존율은 증가함에도 불구하고 유방암 환자의 약 8-10%는 국소 재발을 경험하고 15-30%는 원격 전이를 경험한다. 유방암 재발의 정확한 예측은 초기 개입을 가능하게 하며 생존률을 높이는 것에 기여한다. 기계학습을 이용한 유방암 재발예측 연구가 활발히 진행되고 있는 반면, 국내 자료를 이용한 연구는 아직까지 부족한 실정이다.

본 연구는 2000년 1월 1일부터 2019년 5월 31일까지 신촌 세브란스 병원을 방문하여 유방암을 진단받고 유방절제술을 받은 13,968명 중 진단 시 이미 전이가 있는 환자(120명), 비침윤성 유방암을 진단받은 환자 (2,523명)을 제외한 11,292명을 대상으로 후향적으로 수집한 자료를 이용하였다. 유방암 분자아형에 따른 일반적 특성, 유방암 재발 여부에 따른 일반적 특성은 Chi-square test로 확인하였다. Kaplan-Meier 생존 분석을 사용하여 유방암 위험요인 및 치료변수에 따른 유방암 수술 후 무 재발 생존율을 분석하고 Log rank test로 검정하였다. Cox Proportional Hazard model를 이용하여 유방암 수술 후 암 재발 위험요인을 분석하였다. 유방암 분자아형에 따른 암 재발 위험도가 변수 그룹 별로 차이가 있는지 확인하기 위해 하위그룹 분석을 시행하였다. Cox Proportional Hazard model, Random survival forest, XGBSE로 유방암 재발 예측모형을 구축하고 적합도와 변수 중요도를 비교하였다. 데이터

분석은 R 4.3.4와 Python 3.12.2를 사용하여 분석하였다.

암 재발한 군은 약 11%이며 재발하지 않은 군은 약 89%이다. HR+/HER2-군은 54%으로 가장 많으며 HR+/HER2+군은 18%, HR-/HER2+군은 10%으로 가장 적으며 HR-/HER2-군은 15%이다.

무 재발 생존율은 수술 연도, 진단 시 연령, 유방암 분자 아형, N stage, 내분비 치료 여부, 수술 후 항암치료 여부, 수술 방법에 따라 유의미한 차이를 보였다. 특히, 2010년 이후에 수술받은 환자들의 무 재발 생존율이 2010년 이전에 수술받은 환자들보다 높게 나타났으며($p<0.001$), 30대 이하 환자들 가장 낮은 무 재발 생존율을 보였다($p<0.001$). 유방암 분자 아형 중에서는 삼중 음성 환자들이 가장 낮은 무 재발 생존율을 보였고, T stage, N stage, TNM stage가 진행될수록 무 재발 생존율이 감소하였다($p<0.001$).

콕스 비례위험분석 결과 다른 변수들을 통제하였을 때 진단 시 연령이 높을수록 재발 위험도가 감소하며($p<0.001$), 유방암 분자 아형 HR+/HER2-군 대비 HR+/HER2+군($HR=1.21$, 95% CI =1.02-1.43), HR-/HER2+군($HR=1.53$, 95% CI=1.08-2.15), HR-/HER2-군($HR=2.95$, 95% CI=2.14-4.06)일수록 재발 위험이 증가한다. T stage, N stage 병기가 진행될수록 재발 위험이 증가한다($p<0.001$). 방사선 치료를 받는 경우, 수술 후 항암치료를 받은 경우는 그렇지 않은 경우보다 암 재발 위험이 낮다($p<0.001$). 유방 부분 절제술을 받은 군 대비 유방 전절제를 받은 군의 암 재발 위험이 높으며($p<0.001$), 액와림프절 감시 생검술을 받은 군 대비 액와림프절 광청술을 받은 군의 재발 위험이 높다($p<0.001$).

유방암 분자 아형에 따른 암 재발 상대위험도 하위그룹 분석 결과 2010년 이후 수술받은 경우가 2010년 이전 수술받은 경우보다 유방암 분자 아형이 암

재발에 미치는 영향이 크다($p=0.048$). N stage 2기, 3기인 경우에서 유방암 분자 아형이 암 재발에 미치는 영향이 크다($p=0.018$).

머신러닝 분석 결과 Cox 비례위험 모델보다 머신러닝 모델의 예측력이 낮았다. RSF 모델의 성능은 train set(C-index 0.886)이 Test set(C-index 0.721)보다 좋으며 이는 RSF 모델이 Train set에 과적합 되었다고 볼 수 있다. XGBSE 모델은 Train set(C-index 0.769)과 Test set(C-index 0.692) 모두에서 비슷한 성능을 보이지만 예측력이 가장 낮다.

본 연구는 유방암 분자 아형과 병기에 따른 유방암 재발 위험도를 확인하였고 국내 단일기관 자료를 이용하여 유방암 재발 예측모델을 만들었다는 것에 의의가 있다. 유방암 분자 아형과 병기를 고려하여 환자의 추적관찰 계획을 세워야 한다. 단일기관의 자료만을 이용하였다는 제한점이 있으므로 연구 결과를 기반으로 향후 대규모의 다기관 자료를 이용한 유방암 재발 예측모델에 대한 연구가 필요하다.

핵심어 : 유방암 수술 후 암 재발, 유방암 분자아형, 콕스 비례위험 모형, 머신러닝

I . 서론

1. 연구의 배경 및 필요성

유방암은 전 세계적으로 발생률이 빠르게 증가하고 있으며 우리나라 여성 유방암 발생률이 지난 20년 동안 꾸준히 증가하고 있다. 2000년에는 인구 10만 명당 26.3명이었던 유방암 환자 수가 2010년에는 67.6명, 2015년에는 88.9명으로 증가하였으며, 2019년에는 115.6명으로 더욱 높아져 인구 10만 명 당 100명을 초과하였다. 2020년 암 등록통계에 따르면, 5년 생존율은 93.8%이고 10년 생존율은 89.4%이다(한국유방암학회, 유방암 백서, 2020).

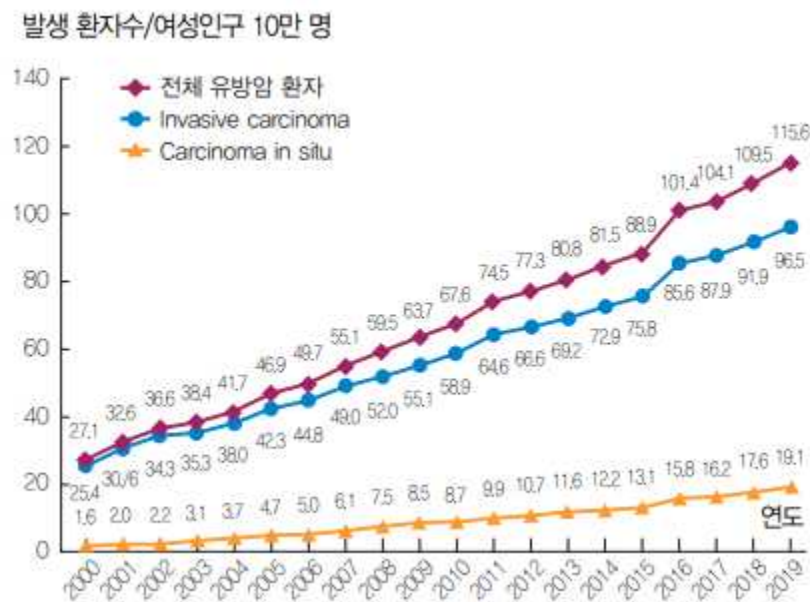


Figure 1. 유방암 증가 현황 (중앙암등록본부, 2021)

조기 발견 및 개선된 의료 기술로 유방암 생존율은 증가함에도 불구하고 유방암 환자의 약 8-10%는 국소 재발을 경험하고 15-30%는 원격 전이를 경험한다. 이전 연구들에 따르면 종양 크기, 겨드랑이 림프절 전이 여부, 유방암 수용체, 방사선 치료 여부, 유방암 진단 시 나이 등이 유방암 재발의 위험요인으로 확인되었다. 특히 삼중 음성 유방암 환자들은 진단 후 처음 3~5년 동안 높은 재발 위험을 갖는다(Pan H et al., 2017).

유방암 재발의 정확한 예측은 초기 개입을 가능하게 하며 전체 생존율을 향상시키는 데 중요하다. 최근 의료 분야를 포함한 여러 분야에서 딥 러닝을 이용한 예측 모델을 개발하고 있다. 복잡한 데이터에서 숨겨진 패턴과 연결을 드러내는 이러한 기술은 유방암 재발 예측의 정밀도와 신뢰성을 증가시킬 수 있다(Clift AK et al., 2023).

기계학습을 이용한 유방암 재발예측 연구가 활발히 진행되고 있는 반면, 국내 자료를 이용한 연구는 아직까지 부족한 실정이다. 본 연구는 단일기관의 코호트 자료를 이용하여 유방암 수술 경험자에서의 유방암 재발 위험요인을 파악하고 기계학습을 이용하여 유방암 재발 예측 모델의 개발 가능성을 확인하고자 한다.

2. 연구의 목적

연구의 목적은 유방암 수술 경험자에게서 발생하는 유방암 재발위험 요인, 유방암 재발예측 모델을 만들어 암의 특성과 환자의 상태에 따른 추적관찰 계획 수립에 기초자료를 제공함에 있다.

이를 달성하기 위한 구체적인 연구 목적은 다음과 같다.

첫째, 연구 대상자인 유방암 수술 경험자의 일반적 특성의 빈도 및 분포를 파악한다.

둘째, 유방암 수술 경험자에서 유방암 분자 아형에 따른 재발 위험을 확인한다.

셋째, 유방암 수술 경험자에서 재발 위험과 관련된 요인을 확인한다.

넷째, 유방암 수술 경험자에서 재발 예측 모델을 구축하고 적합도를 확인한다.

II. 이론적 배경

1. 유방암 원인 및 치료법

유방암은 유방에 발생한 암세포로 이루어진 종괴로 유방의 유관과 유엽에서 발생하는 암을 의미한다. 주로 침윤성 유방암(7-80%), 비 침윤성 유방암(5-15%), 그리고 특수 유형의 암 등으로 구분한다. 침윤성 유방암은 암세포가 유방의 다른 부위로 침입하거나 확산할 수 있어 더 진행된 형태로, 유방 외의 다른 신체 부위로 전이될 가능성이 있다. 비침윤성 유방암은 암세포가 유방의 원래 위치에서 벗어나지 않고 유방의 특정 조직 내에서만 자라난다. Ductal Carcinoma In Situ (DCIS)는 가장 흔한 비침윤성 유방암 유형으로, 유방의 관(ducts) 내에서 발생한다. Lobular Carcinoma In Situ (LCIS)은 유방의 소엽(lobules)에서 발생하며, DCIS보다는 암으로 발전할 위험이 낮다. 비침윤성 유방암은 일반적으로 국소적으로 치료가 가능하지만, 침윤성 유방암은 종종 수술, 방사선 치료, 화학 요법 및 호르몬 치료와 같은 보다 복합적인 접근이 필요하다(NCCN, 2018).

유방암의 발생 원인이 명확하게 규명된 것은 아니지만 유방암의 위험 요인으로 이른 초경, 늦은 폐경, 임신 경험이 없는 것과 늦은 연령의 첫 만삭 임신, 폐경 후 비만 등의 내부 요인이 있다. 이 외에도, 음주와 흡연 등의 생활습관, 유방암 가족력과 BRCA1/2 유전자 돌연변이 등 개인적 소인과 관련된 요인, 치밀유방, 유방 증식 병변의 과거력 등, 경구피임약 복용 등이 있다(Barke LD, 2017).

유방암의 치료는 발생 연령, 침윤 정도, 암종의 병리학적 특성, 암조직의 호르몬수용체(estrogen receptor, progesterone receptor)의 유무 및 성장요인 수용체, c-erb-B-2등 epidermal growth factor receptor (EGFR) families 유무, 등을 고려하여 결정한다. 유방암의 치료법으로 국소적 요법으로는 수술요법, 방사선요법이 있으며, 전신요법으로는 화학요법, 면역호르몬요법, 표적 치료가 있고 이들을 2가지 이상 같이 시행하는 다학제적 병용요법 등이 있다. 이때 어떤 치료법들을 조합하여 결정하며 이에 따라 재발률, 생존율이 다를 수 있다(Paik, 2014).

유방암은 호르몬 수용체와 HER2에 발현양상에 따라 분자학적으로 분류한다. 호르몬 수용체 양성/HER2 음성 유방암은 전체 유방암의 약 74%를 차지한다. 천천히 증식하며 항암 내분비 요법에 잘 반응하여 예후가 좋다. 호르몬 수용체 양성/ HER2양성 유방암은 전체 유방암의 약 10%를 차지한다. 호르몬 수용체 음성/HER2양성 유방암은 전체 유방암의 약 4%를 차지하며 공격적이며 단기간의 예후가 좋지 않다. 호르몬 수용체 음성/ HER2 음성 유방암은 전체 유방암의 약 12%를 차지하며 예후가 가장 좋지 않다(한국 유방암 학회, 2022).

2. 기계학습

머신러닝은 컴퓨터가 데이터를 통해 학습하고 경험을 통해 개선하도록 훈련하는 인공지능(AI)의 하위 집합이다. 머신러닝에서 알고리즘은 대규모 데이터 세트에서 패턴과 상관관계를 찾고 분석을 토대로 최적의 의사결정과 예측을 수행하도록 훈련된다. 머신러닝 기술은 대량의 환자 의료 기록이나 데이터를 처리할 수 있으며, 이는 생존 예측의 정확성과 신뢰성을 향상시킨다. 딥러닝은 인공 신경망을 사용하여 고차원 데이터에서 패턴을 추출하고 예측을 수행하는 새로운 머신러닝 기술이다. 이는 초기 암 탐지에서 좋은 성능을 보이며 헬스케어 분야에서 다수 활용되고 있다(Tran, 2021).

랜덤 포레스트는 배깅의 대표적인 방법으로, 의사결정나무의 과적합을 해결하는 학습모형이다. 배깅을 통해 랜덤 포레스트를 구축하는 방법은 다음과 같다. B개의 부트스트랩(bootstrap) 표본집합을 생성한다. 각 표본집합을 통해 B개의 의사결정나무를 훈련시킨다. B개의 의사결정나무를 하나의 랜덤 포레스트로 결합 후, 최종적으로 과반수 원칙에 따라 가장 많이 분류된 범주로 반응변수를 예측한다. 이와 같이 형성된 랜덤 포레스트는 각각의 의사결정나무의 노드마다 분리에 사용할 후보 변수들을 무작위로 선택하고, 가지치기를 수행하지 않기 때문에 예측한 반응변수에 대해 분산과 오차가 낮다는 장점이 있다. 따라서 다양한 분석에 랜덤 포레스트가 사용되어왔고, 생존 분석에도 접목되기 시작했다(Ishwaran et al., 2008).

랜덤 서바이벌 포레스트(Random Survival Forests, RSF)는 머신러닝의 트리 기반 모델인 랜덤 포레스트(Random Forest, Breiman, 2001)의 확장이다. 전통적인 생존 분석 방법들이 모수적이거나 특정 가정에 의존하는 반면, RSF는 CIF를 직접 추정하여 정확한 예측 성능을 제공한다. 비선형 효과와 상호작용을 모델링하는 장점을 갖기 때문에 고차원 설정에서 효과적으로 사용할 수 있다. 기존의 생존 분석에는 콕스의 비례위험모형은 비례 위험 가정을 만족해야 하거나 모수에 대한 사전 지식을 필요로 하지만, 랜덤 포레스트를 접목시킴으로써 이러한 문제점을 해결할 수 있다 (Ishwaran et al., 2008).

XGBSE (Extreme Gradient Boosting Survival Estimation)는 XGBoost를 기반으로 한 생존 분석 방법이다. XGBoost는 그라디언트 부스팅을 사용하는 머신러닝 알고리즘으로, 의사결정 나무가 Base Learner이다. 잔차를 이용하여 이전 모형의 약점을 보완하는 방식으로 강력한 예측 모델을 생성한다. XGBSE는 이를 확장하여 생존 데이터에 적합한 모델을 구축하며, 생존 시간 추정과 이벤트 시점의 확률을 예측하는 데 사용한다(Chen, 2016).

하렐의 C-index는 생존 분석에서 예측 오류를 측정하는 데 일반적으로 사용되는 지표로, 높은 위험 점수를 생성하는 모델을 비교하는 데 사용된다. 이는 검열된 데이터를 고려하는 ROC 곡선 아래의 면적(AUC)의 일반화된 형태이다(Harrell JR et al., 1996). 이는 예측대로 더 오래 살 것으로 예상되었던 환자가 실제로는 더 일찍 사망한 경우의 비율을 나타낸다(Jeremy et al., 2011). 무작위로 선택된 환자들 사이에서, 사건 발생 시간이 더 짧은 환자는 더 높은 위험 점수와 더 낮은 예측된 사건 발생 시간 결과를 가지게 된다. C-index 값은 0과 1 사이의 비율로 나타난다. 1에 가까운 값은

모델의 높은 구별 성능을 나타내며, 0.5에 가까운 값은 무작위 예측과의 유사성을 보여준다(Harrell et al., 1996).

Ⅲ. 연구 방법

1. 연구 모형

<Figure 2>은 본 연구의 틀이다.

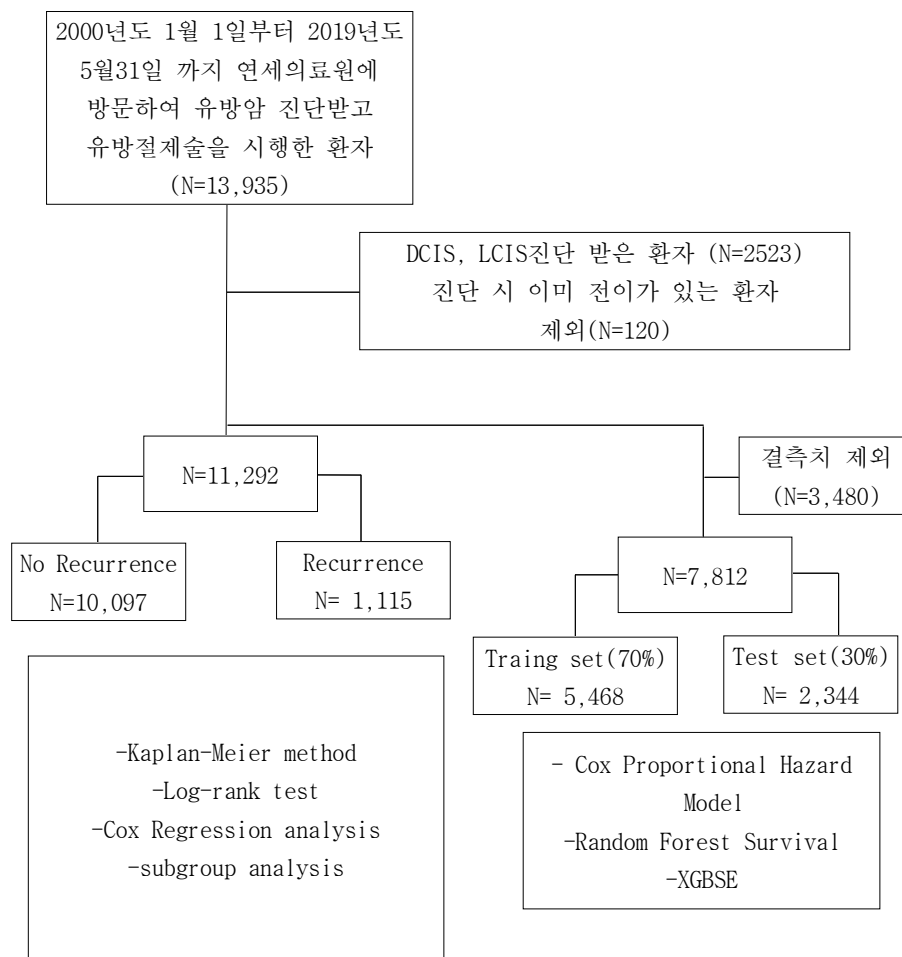


Figure 2. Framework of study

2. 연구 대상 및 자료원

본 연구는 2000년 1월 1일부터 2019년 5월 31일까지 신촌 세브란스 병원을 방문하여 유방암을 진단받고 유방절제술을 받은 13,935명을 대상으로 병리학 적 기록을 후향적으로 수집한 자료를 이용하였다. DCIS, LCIS진단 받은 환자 2,523명을 제외하였으며 진단 시 이미 원격 전이가 있는 환자 120명을 제외 하여 최종 11,292명을 대상으로 하였다.

3. 변수의 선정 및 정의

가. 종속변수

1) 유방암 재발

종속변수는 유방암 재발이다. 유방암 재발은 국소 재발, 구역 재발, 원격 재발 모두를 포함한다.

2) Recurrence Free Survival

유방암 재발이 있는 경우, 수술 시행일로부터 유방암 재발 날짜까지의 기간을 RFS로 정의하였다. 관찰 기간 동안 유방암 재발이 발생하지 않은 경우, 마지막 관찰 시점까지의 기간을 RFS로 정의하였다. 유방암 재발 없이 사망한 경우, 재발(event) 처리하였고 수술 시행일로부터 사망 시점까지의 기간을 RFS로 정의하였다. 본 연구에서 최대 관찰 기간은 10년으로 설정하였다.

나. 독립변수

1) BMI Index

BMI는 체중을 키의 제곱으로 나눈 값으로 비만도를 분류하는 데 일반적으로 사용되는 지표이다. 세계보건기구(WHO)에 따르면 25kg/m^2 미만은 정상, 25kg/m^2 이상은 과체중, 30kg/m^2 이상은 비만으로 정의한다.

2) 유방암 분자 아형

호르몬 수용체(HR)는 면역조직화학(IHC) 염색 검사에서 에스트로겐 또는 프로게스테론 수용체로 염색된 세포가 1 이상인 경우로 정의되었다. HER2 상태는 면역조직화학(IHC) 또는 FISH 검사를 통해 평가한다. IHC 결과 0 또는 1+는 HER2 음성, 3+는 HER2 양성으로 정의한다. FISH결과 HER2 유전자와 세포질 내 크로모솜 17의 비율이 2.0 이상일 경우 HER2 양성 정의한다(Partin JF, 2011).

호르몬 수용체 양성,HER2 음성 유방암을 HR+/HER2-로 표기하였고, 호르몬 수용체 양성,HER2 양성 유방암을 HR+/HER2+로 표기 하였다. 호르몬 수용체 음성,HER2 양성 유방암을 HR-/HER2+로 표기하였고, 호르몬 수용체 음성,HER2 음성 유방암을 HR-/HER2-로 표기 하였다.

3) 암 가족력

가족, 친척 중에 유방암을 포함한 모든 암에 대해 암 환자가 있는 경우를 ‘가족력이 있음’ 으로 정의하였다. 가족, 친척 중에 유방암을 포함한 모든 암에 대해 암 환자가 없는 경우를 ‘가족력이 없음’ 으로 정의하였다.

4) T stage

T stage는 종양의 크기와 특성에 따라 다음과 같이 분류한다. 종양의 증거가 없는 경우 T0로 분류한다. 관상피내암 또는 소엽상피내암인 경우 Tis로 분류한다. 종양의 최대 직경이 2cm 이하이면 T1, 2cm초과 5cm이하인 경우 T2, 5cm초과하는 경우 T3로 분류한다. 종양이 흉벽이나 피부를 침범한 경우는 T4로 분류한다(AJCC, 2018).

5) N stage

N stage는 액와 림프절 전이 정도에 따라 분류한다. 림프절 전이가 없는 경우 N0로 분류한다. 전이된 림프절의 개수가 3개 이하인 경우는 N1으로 분류한다. 4개 이상 9개 이하인 경우 N3로 분류하며, 10개 이상인 경우 N3으로 분류한다(AJCC, 2018).

6) TNM Stage

유방암이 세포 내에 국한 되어 있는 상피 내암인 경우 0기로 분류한다. 암의 크기가 2cm 이하이며 림프절 전이와 원격전이가 없는 경우 1기로 분류한다. 종양의 크기가 2cm초과 5cm이하 이면서 림프절 전이 개수가 3개 이하이며 원격전이가 없는 경우 2기로 분류한다. 암의 크기가 5cm이상이면서 림프절 전이가 있고 원격전이가 없는 경우 3기로 분류한다(AJCC, 2018).

7) Endocrine Therapy

내분비 요법은 호르몬 수용체 양성 유방암 환자에게 적용되는 치료법이다. 호르몬 수용체 양성 유방암에 내분비 요법을 통하여 호르몬 활동을 억제하거나 방해함으로 암세포의 성장을 늦춘다(Paik, 2014).

8) Radiation Therapy

방사선 치료는 고에너지 입자와 파동을 이용해 암세포를 제거하는 치료이다(Paik, 2003). 수술 후 방사선 치료를 시행한 경우로 정의하였다.

9) Postoperaraive chemo Therapy

수술 후 항암화학요법은 수술 후 CMF, FAC, AC, AT 등의 화학요법이 있으며 유방암의 유형, 병기에 따라 결정한다(Paik, 2014). 수술 후 항암화학요법을 시행한 것으로 정의하였다.

10) 유방암 수술 종류

유방암이 있는 유방의 전체를 절제하는 경우 유방 전절제(Total Mastectomy), 유방암이 있는 조직과 주변의 정상조직의 일부만 절제하는 경우 유방 부분 절제(Partial Mastectomy)라고 정의한다(Paik, 2014).

임상적으로 액와 림프절 전이가 의심되지 않는 유방암 환자에서 감시 림프절 생검술(SLNB)을 시행하며, 감시 림프절 생검술 (SLNB)에서 림프절 전이가 있는 경우에만 액와 림프절 광청술 (ALND)을 시행한다(Lyman GH, 2005). 감시 림프절 침범 여부에 따라 감시림프절만 절제하는 경우 Sentinel Lymph node Boipsy로 정의하였고 겨드랑이 림프절 광청술을 시행하는 경우 Axillary Lymph node Dissection으로 정의하였다.

11) Axillary Lymph node positive 여부

액와 림프절 전이 여부는 유방암 수술의 범위를 정하고 전신 항암화학요법 및 방사선 치료 여부를 결정하는 것에 중요한 지표가 된다(Beenken SW, 2003). 액와 림프절에서 암세포가 발견된 것을 Axillary Lymph node Positive로 정의 한다.

4. 분석방법

유방암 분자 아형에 따른 일반적 특성, 유방암 재발 여부에 따른 일반적 특성은 Chi-square test로 확인하였다.

Kaplan-Meier curve를 사용하여 유방암 위험요인 및 치료 변수에 따른 유방암 수술 후 암 재발률의 시각적 차이를 확인하고 유의한 차이가 있는지 Log rank test를 실시하였다. Cox Proportional Hazard model를 이용하여 변수별 암 재발 비례위험비를 분석하였다. 유방암 분자아형에 따른 암 재발 위험도가 변수그룹 별로 차이가 있는지 확인하기 위해 하위그룹 분석을 시행하였다.

유방암 재발 예측모형은 Cox Proportional Hazard model, Random survival forest와 XGBSE를 사용하였다. 내적 타당도 검증을 위해 전체 대상자의 70를 훈련용 데이터, 30를 테스트용 데이터로 나누어 분석하였다.

데이터 분석은 R 4.3.4와 Python 3.12.2를 사용하여 분석하였다. 데이터 전처리, 분석 및 시각화를 위해 Python의 Pandas, XGBSE, NumPy 및 Matplotlib 라이브러리를 사용하였다.

모든 분석에서 통계적 유의성을 평가하기 위해 양측 검정을 실시하였으며, p-value가 0.05 미만을 통계적으로 유의한 것으로 정의하였다.

5. 윤리적 고려사항

연구 대상자의 권리와 안전을 보호하기 위해 세브란스의 기관생명윤리위원회(Institutional Review Board, IRB)의 승인을 받았습니다(IRB 승인번호: 2023-3336-001). 본 연구는 후향적 자료를 이용한 연구로, 연구 참여자들로 부터 별도의 동의 절차는 필요하지 않다. 모든 연구 절차는 IRB의 윤리적 가이드라인을 준수하였으며, 수집된 자료는 익명으로 처리되어 개인 식별이 불가능하도록 관리하였다.

IV. 연구 결과

1. 연구 대상자의 일반적 특성

1.1. 유방암 재발 여부에 대한 일반적 특성

총 연구 대상자는 11,292명이며 암 재발한 군은 1115명으로 약 11%이며 재발하지 않은 군은 10,097명으로 약 89%이다. 유방암 재발 여부에 따른 일반적 특성을 카이 스퀘어 검정으로 분석하였고 연속형 변수는 t-test 분석하였다.

통계적으로 유의한 변수는 다음과 같다. 암 재발 군의 진단 시 연령 평균(49.05세)은 암 비 재발 군의 연령 평균(51.15세)보다 낮았다($p<0.001$). 수술 받은 연도가 2000-2009인 경우의 암 비 재발군(84.3)의 분율은 2010년이 후 수술 받은 경우의 암 비 재발 군(92.8)보다 작다($p<0.001$). 암 재발군의 분율은 BMI는 과체중군, 비만군 순으로 크다($p<0.001$). 연령에서는 30대 이하의 암 재발군의 분율(17.3)이 가장 크고 연령이 높아질수록 낮아진다($p=0.02$). 유방암 분자 아형에서는 HR-/HER2-에서의 암 재발군의 분율이 15.5%로 가장 컸으며 HR+/HER2-에서 7.5%로 가장 작았다($p<0.001$). T stage, N stage, TNM stage에서 병기가 진행될수록 암 재발군의 분율이 증가하였으며 T stage 4기에서 52.4%, N stage 3기에서 50.3%, TNM stage 4기에서 52.4%로 가장 컸다($p<0.001$).

치료 변수에 대한 유의미한 결과는 다음과 같다. 수술 후 내분비 치료를 받

은 경우 암 재발군의 분율(13.29)은 항암치료 받지 않은 경우의 암 재발군 분율(8.67)보다 크다($p<0.001$). 수술 후 항암치료를 받은 군의 암 재발군 분율(12.7)은 항암치료 받지 않은 군(6.9)보다 크다($p<0.001$). 유방 절제술 방법에서 전절제를 받은 경우의 암 재발군 분율(17.9)은 부분 절제술(5.9)보다 크다($p<0.001$). 액와 림프절 수술 방법에서 액와 림프절 광청술을 받은 경우의 암 재발군 분율(17.9)은 액와 감시림프절 생검술을 받은 경우(5.5)보다 크다($p<0.001$). 액와 림프절 양성인 경우의 암 재발군 분율(19.0)은 음성인 경우(6.1)보다 크다($p<0.001$).

성별, 모유 수유 경험 여부, 출산 경험 여부, 가족력, 내분비 치료 여부, 방사선 치료 여부는 통계적으로 유의한 차이를 보이지 않았다.

Table.1 Baseline characteristics by Breast cancer recurrence.

Variables	Patients, by Recurrences		p-value
	No Recurrence (n=10097)	Recurrence (n=1115)	
Age(mean)	51.15	49.05	<0.001
sex			1
Female	9597(89.6)	1112(10.4)	
male	30(90.9)	3(9.1)	
Year of operation			<0.001
2000-2009	3062(84.3)	569(15.7)	
2010-2019	7035(92.8)	546(7.2)	
BMI			<0.001
<25	6285(90.9)	627(9.1)	
<30	3026(89.2)	367(10.8)	
≥ 30	549(90.4)	58(9.6)	
Age			0.02
<40	1182(82.7)	247(17.3)	
<50	3596(91.0)	358(9.1)	
<60	3004(90.8)	303(9.2)	
≥ 60	2314(91.8)	206(8.2)	
Breast feeding			0.76
no	3346(90.8)	338(9.2)	
yes	6090(90.6)	630(9.4)	
Parity			0.17
no	567(88.6)	73(11.4)	
yes	8038(90.3)	860(9.7)	
Presence of family history of cancer			0.09
no	5163(89.5)	609(10.6)	
yes	4079(90.5)	429(9.5)	
Subtype			<0.001
HR+/HER2-	5524(92.5)	448(7.5)	
HR+/HER2+	1768(88.4)	233(11.6)	
HR-/HER2+	1024(88.3)	136(11.7)	
HR-/HER2-	1462(84.5)	269(15.5)	
T stage			<0.001
T0/Tis	673(94.79)	37(5.21)	
I	6881(92.8)	536(7.2)	
II	2398(84.1)	453(15.9)	
III	111(64.2)	62(35.8)	
IV	20(47.6)	22(52.4)	

Table.1 Baseline characteristics by Breast cancer recurrence(continued).

Variables	Patients, by Recurrences		p-value
	No Recurrence (n=10097)	Recurrence (n=1115)	
N stage			<0.001
0	7379(93.9)	477(6.1)	
I	2017(86.5)	316(13.5)	
II	476(75.7)	153(24.3)	
III	155(49.7)	157(50.3)	
TNM Stage			<0.001
0	637(95.07)	33(4.93)	
I	5487(95.11)	282(4.89)	
II	3296(87.68)	463(12.32)	
III	677(66.77)	337(33.23)	
Endotherapy			<0.001
None	2695(86.71)	413(13.29)	
Done	7396(91.33)	702(8.67)	
Radiation therapy			1
None	3399(90.0)	376(10.0)	
Done	6691(90.1)	739(9.9)	
Postoperative			<0.001
Chemotherapy			
None	4924(93.1)	363(6.9)	
Done	5163(87.3)	752(12.7)	
Breast operation			<0.001
PM	5568(94.1)	349(5.9)	
TM	4512(85.5)	763(14.5)	
Axillary operation			<0.001
SLNB	6766(94.5)	391(5.5)	
ALND	3321(82.1)	723(17.9)	
Axillary Lymphnode			<0.001
negative	7306(93.9)	471(6.1)	
positive	2680(81.0)	628(19.0)	

****PM: Partial Mastectomy, TM: Total Mastectomy**

1.2. 유방암 분자아형에 따른 일반적 특성

유방암 분자 아형에 따른 일반적 특성의 범주형 변수는 카이스퀘어 검정을 이용하여 분석하였고 연속형 변수는 Anova검정을 시행하였다. HR+/HR-군은 5977명(54%)으로 가장 많으며 HR+/HER2+군은 2003명(18%), HR-/HER2+군은 1163명(10%)으로 가장 적으며 HR-/HER2-군은 1735명(15%)이다.

통계적으로 유의한 차이를 보이는 변수는 다음과 같다. 진단 시 연령은 HR+/HER2+군(50.1세)이 가장 작고 HR-/HER2+군(52.3세)이 가장 크다($p<0.001$). 유방암 분자아형 4군 모두 수술 받은 연도가 2010년 이후인 경우의 분율이 2010년도 이전인 경우보다 크다($p<0.001$). 연령에서 40대가 가장 많은 군은 HR+/HR-군(38.4%)과 HR+/HER2+군(37.2%)이었으며 50대가 가장 많은 군은 HR-/HER2+군(40.2%) HR-/HER2-군(30.1%)이었다($p<0.001$). 모유 수유 경험에서는 네 군 모두 모유 수유 경험 있는 경우의 분율이 가장 크다($p<0.01$). 가족력에서는 네 군 모두 가족력 유무의 분율이 비슷했다($p<0.001$). T stage에서 1기의 분율이 호르몬 양성인 HR+/HR-군(72.9%)과 HR+/HER2+군(65.8%)이 호르몬 음성 HR-/HER2+군(52.7%)과 HR-/HER2-군(53.9%)보다 컸으며, 2기의 분율은 호르몬 음성 HR-/HER2+군(29.1%)과 HR-/HER2-군(29.7%)이 호르몬 양성 HR+/HR-군(23.5%)과 HR+/HER2+군(28.3%)보다 크다($p<0.001$). N stage에서 4군 모두 0기, 1기 순으로 가장 많이 분포하며 HR-/HER2-군에서의 분율(77.4%)이 가장 크다($p<0.001$). N stage 1기의 분율은 호르몬 양성인 HR+/HR-군(22.8%)과 HR+/HER2+군(22.8%)이 호르몬 음성인 HR+/HR-군(19.8%)과 HR+/HER2+군(15.6%)보다 크다($p<0.001$). TNM stage에서 네 군 모두 1기, 2기 순으로 많았다($p<0.001$).

치료변수에 대하여 유의미한 결과는 다음과 같다. 호르몬 양성 유방에 주로 적용되는 내분비 치료에서 치료 받은 경우의 분율은 호르몬 양성인 HR+/HR-군(96.6%)과 HR+/HER2+군(96.6%)이 호르몬 음성인 HR-/HER2+군(6.1%)과 HR-/HER2-군(5.0%)보다 크다($p<0.001$). 네 군 모두 방사선 치료를 받은 경우의 분율이 시행하지 않은 군보다 크다($p<0.001$). 수술 후 항암치료 여부에서 HR+/HR-군을 제외한 세 군은 수술 후 항암치료를 받지 않은 경우가 더 많았고 HR+/HR-군은 수술 후 항암치료를 받은 경우(57.6%)가 더 많았다($p<0.001$). 유방 절제술 종류의 경우 HER2 양성인 HR-/HER2+군(61.4%)과 HR+/HER2+군(53.2%)은 전절제술을 받은 경우가 많으며 HER2 음성인 HR+/HER2-군(56.0%)과 HR-/HER2-군(58.2%)은 부분절제술을 받은 경우가 많다($p<0.001$). 액와림프절 수술에서는 네 군 모두 감시림프절 생검술을 시행한 경우가 액와림프절 광청술을 받은 경우보다 많았고, HR-/HER2-군에서 감시림프절 생검을 받은 분율(65.8%)이 가장 크다($p<0.001$). 액와림프절 전이에 대해서는 네 군 모두 음성인 경우가 양성인 경우보다 많았으며 HR-/HER2-군의 액와림프절 음성인 경우의 분율(77.3%)이 가장 크다($p<0.001$).

유방암 분자 아형에 따른 성별, 체질량 지수, 출산 경험 여부의 차이는 통계적으로 유의하지 않다.

Table.2 Baseline characteristics by Breast cancer subtype.

Variables	Patients, by Breast cancer Subtype				P-value
	HR+/HER2- (n=5977)	HR+/HER2+ (n=2003)	HR-/HER2+ (n=1163)	HR-/HER2- (n=1735)	
age(mean)	51.6	50.1	52.3	50.6	<0.001
sex					0.17
Female	5953	1194	1162	1732	
male	24	9	1	3	
Year of operation					<0.001
2000-2009	1717 (28.7)	807 (40.3)	440 (37.8)	630 (36.3)	
2010-2019	4260 (71.3)	1196 (59.7)	723 (62.2)	1105 (63.7)	
BMI					0.14
<25	3685 (62.8)	1228 (64.0)	723 (64.2)	1065 (63.7)	
<30	1818 (31.0)	598 (31.2)	355 (31.5)	517 (30.9)	
≥ 30	364 (6.2)	93 (4.8)	49 (4.3)	90 (5.4)	
Age					<0.001
<40	636 (10.6)	282 (14.1)	141 (12.1)	333 (19.2)	
40-49	2293 (38.4)	745 (37.2)	291 (25.0)	491 (28.3)	
50-59	1627 (27.2)	609 (30.4)	467 (40.2)	522 (30.1)	
≥ 60	1419 (23.7)	367 (18.3)	264 (22.7)	389 (22.4)	
Breast feeding					<0.01
no	2028 (35.8)	655 (36.3)	300 (28.7)	586 (37.0)	
yes	3632 (64.2)	1147 (63.7)	744 (71.3)	996 (63.0)	
Parity					0.17
no	361 (7.0)	95 (5.7)	50 (4.9)	117 (7.9)	
yes	4766 (93.0)	1566 (94.3)	967 (95.1)	1364 (92.1)	
Presence of family history of cancer					<0.001
no	3025 (55.5)	1023 (55.4)	634 (59.0)	906 (56.6)	
yes	2424 (44.5)	824 (44.6)	440 (41.0)	696 (43.4)	
T stage					<0.001
T0/Tis	141(2.36)	87(4.35)	174(15)	221(12.77)	
I	4356 (72.9)	1316 (65.8)	611 (52.7)	933 (53.9)	
II	1402 (23.5)	566 (28.3)	338 (29.1)	514 (29.7)	
III	61 (1.0)	21 (1.0)	28 (2.4)	54 (3.1)	
IV	13 (0.2)	10 (0.5)	9 (0.8)	9 (0.5)	

Table.2 Baseline characteristics by Breast cancer subtype(continued).

Variables	Patients, by Breast cancer Subtype				P-value
	HR+/HER2- (n=5977)	HR+/HER2+ (n=2003)	HR-/HER2+ (n=1163)	HR-/HER2- (n=1735)	
N stage					<0.001
0	4123 (69.3)	1329 (66.5)	809 (69.7)	1338 (77.4)	
I	1357 (22.8)	455 (22.8)	230 (19.8)	269 (15.6)	
II	325 (5.5)	145 (7.3)	76 (6.5)	72 (4.2)	
III	147 (2.5)	69 (3.5)	46 (4.0)	49 (2.8)	
TNM Stage					<0.001
0	137 (2.3)	72 (3.6)	160 (13.8)	211 (12.2)	
I	3387 (56.7)	980 (48.9)	462 (39.7)	768 (44.3)	
II	1956 (32.7)	724 (36.1)	409 (35.2)	615 (35.4)	
III	497 (8.3)	227 (11.3)	132 (11.3)	141 (8.1)	
Endotherapy					<0.001
None	201 (3.4)	68 (3.4)	1091 (93.9)	1648 (95.0)	
Done	5774 (96.6)	1933 (96.6)	71 (6.1)	87 (5.0)	
Radiation therapy					<0.001
None	1969 (33.0)	716 (35.8)	459 (39.5)	464 (26.7)	
Done	4006 (67.0)	1285 (64.2)	702 (60.5)	1271 (73.3)	
Postoperative					<0.001
Chemotherapy					
None	3440 (57.6)	667 (33.4)	340 (29.3)	596 (34.4)	
Done	2534 (42.4)	1331 (66.6)	822 (70.7)	1139 (65.6)	
Breast operation					<0.001
PM	3344 (56.0)	935 (46.8)	447 (38.6)	1008 (58.2)	
TM	2628 (44.0)	1065 (53.2)	712 (61.4)	723 (41.8)	
Axillary operation					<0.001
SLNB	3918 (65.6)	1193 (59.6)	685 (58.9)	1141 (65.8)	
ALND	2054 (34.4)	810 (40.4)	478 (41.1)	592 (34.2)	
Axillary Lymph node					<0.001
negative	4085 (68.8)	1321 (66.1)	806 (69.5)	1330 (77.3)	
positive	1850 (31.2)	677 (33.9)	353 (30.5)	391 (22.7)	

****PM: Partial Mastectomy, TM: Total Mastectomy**

2. 유방암 수술 후 암 재발률 및 재발 비례위험도

2.1. 유방암 수술 후 무 재발 생존에 영향을 미치는 요인

Kaplan-Meier test를 사용하여 유방암 위험요인 및 치료 변수들에 대하여 유방암 수술 후 무 재발 생존율(Recurrence Free Survival)을 계산하고 비례 위험 가정을 만족하는지 확인하였다. log rank test로 생존곡선이 통계적으로 유의한 차이가 있는지 확인하였다.

무 재발 생존율의 통계적으로 유의미한 결과는 다음과 같다. 수술을 시행한 연도가 2010년 이전인 경우의 3년(0.91), 5년(0.88), 10년(0.85) 무 재발 생존율은 2010년 이후인 경우의 3년(0.95), 5년(0.91), 10년(0.85) 무재발 생존율보다 낮다($p < 0.001$). 연령에서 30대 이하인 경우의 3년(0.89), 5년(0.84), 10년(0.75)의 무 재발 생존율이 가장 낮았다($p < 0.001$). 유방암 분자아형에서 HR+/HER2-군의 경우 3년(0.96), 5년(0.93), 10년(0.85)의 무재발 생존율이 가장 높고, 삼중 음성인 경우의 3년(0.87), 5년(0.83), 10년(0.79)의 무 재발 생존율이 가장 낮았다($p < 0.001$). T stage에서 병기가 진행될수록 무재발 생존율이 감소하며 T stage 4기인 경우의 3년(0.44), 5년(0.37), 10년(0.37)의 무 재발 생존율이 가장 낮았다($p < 0.0001$). N stage에서 stage가 증가할수록 무 재발 생존율이 감소하고, N stage가 3기인 경우 3년(0.61), 5년(0.50), 10년(0.36)의 무 재발 생존율이 가장 낮았다($p < 0.001$). TNM stage에서 stage가 증가할수록 무 재발 생존율이 감소하고, 3기인 경우 3년(0.76), 5년(0.67), 10년(0.55) 무 재발 생존율이 가장 낮았다($p < 0.001$). 내분비 치료를 받은 환자의 3년(0.96), 5년(0.92), 10년(0.84) 무 재발 생존

율이 내분비 치료를 받지 않은 환자의 3년(0.88), 5년(0.86), 10년(0.82) 무 재발 생존율보다 높다($p < 0.001$). 방사선 치료를 받은 환자의 3년(0.93), 5년(0.90), 10년(0.84) 무 재발 생존율은 방사선 치료를 받지 않은 환자의 3년(0.94), 5년(0.90), 10년(0.84) 무 재발 생존율과 비슷하다($p < 0.001$). 수술 후 항암치료를 받은 환자의 3년(0.92), 5년(0.89), 10년(0.82)의 무 재발 생존율은 수술 후 항암치료를 받지 않은 환자의 3년(0.95), 5년(0.92), 10년(0.86)의 무 재발 생존율보다 낮다($p < 0.001$). 유방 전 절제술을 받은 환자의 3년(0.90), 5년(0.86), 10년(0.79)의 무 재발 생존율은 유방 부분 절제술을 받은 환자의 3년(0.97), 5년(0.94), 10년(0.89)의 무 재발 생존율이 낮다($p < 0.001$). 액와 감시 림프절 생검술을 받은 환자의 3년(0.97), 5년(0.94), 10년(0.90) 무 재발 생존율은 액와 림프절 광청술을 받은 환자의 3년(0.88), 5년(0.83), 10년(0.74) 무 재발 생존율보다 높다($p < 0.001$). 액와 림프절 양성인 환자의 3년(0.88), 5년(0.82), 10년(0.72) 무 재발 생존율은 액와 림프절 음성인 환자의 3년(0.96), 5년(0.94), 10년(0.89) 무 재발 생존율보다 낮다($p < 0.001$).

모유수유 경험 (Log rank test, $p=0.5$), 출산 경험 (Log rank test, $p=0.22$), 가족력(Log rank test, $p=0.65$), 방사선 치료 여부(Log rank test, $p=0.92$)는 유의확률 0.05를 초과하여 수술 후 재발률에 통계적으로 유의미한 차이를 확인할 수 없었다.

모든 변수들의 Kaplan-Meier Curve에서 그래프가 교차되는 지점이 없음을 확인하였고 비례 위험 가정을 만족함을 알 수 있었다.

Table. 3 Recurrence-Free Survival Rates of 3-Year, 5-Year, 10-Year.

Variables	Recurrence Free Survival Rate			P-value
	3 yr	5 yr	10 yr	
Year of operation				<0.001
2000-2009	0.91	0.88	0.85	
2010-2019	0.95	0.91	0.85	
BMI				<0.001
<25	0.94	0.91	0.85	
<30	0.93	0.89	0.82	
≤30	0.94	0.90	0.83	
Age				<0.001
<40	0.89	0.84	0.75	
<50	0.95	0.92	0.85	
<60	0.94	0.91	0.85	
≥60	0.94	0.90	0.85	
Breast feeding				0.02
no	0.94	0.91	0.85	
yes	0.94	0.90	0.84	
Parity				0.2
no	0.93	0.89	0.82	
yes	0.94	0.90	0.84	
Presence of family history of cancer				<0.01
no	0.93	0.90	0.83	
yes	0.94	0.91	0.84	
Subtype				<0.0001
HR+/HER2-	0.96	0.93	0.85	
HR+/HER2+	0.94	0.90	0.83	
HR+/HER2-	0.89	0.87	0.85	
HR2-HER2-	0.87	0.83	0.79	
T stage				<0.0001
T0/Tis	0.95	0.92	0.83	
I	0.96	0.93	0.87	
II	0.89	0.84	0.77	
III	0.67	0.60	0.55	
IV	0.44	0.37	0.37	

Table. 3 Recurrence-Free Survival Rates of 3-Year, 5-Year, 10-Year (continued).

Variables	Recurrence Free Survival Rate			P-value
	3 yr	5 yr	10 yr	
N stage				<0.0001
0	0.96	0.94	0.89	
I	0.92	0.87	0.79	
II	0.85	0.77	0.66	
III	0.61	0.50	0.36	
TNM Stage				<0.0001
0	0.97	0.94	0.88	
I	0.97	0.95	0.91	
II	0.92	0.88	0.82	
III	0.76	0.67	0.55	
Endotherapy				<0.001
None	0.88	0.86	0.82	
Done	0.96	0.92	0.84	
Radiation therapy				<0.001
None	0.94	0.90	0.84	
Done	0.93	0.90	0.84	
Postoperative Chemotherapy				0.01
None	0.95	0.92	0.86	
Done	0.92	0.89	0.82	
Breast operation				<0.0001
PM	0.97	0.94	0.89	
TM	0.90	0.86	0.79	
Axillary operation				<0.0001
SLNB	0.97	0.94	0.90	
ALND	0.88	0.83	0.74	
Axillary Lymph node				<0.0001
negative	0.96	0.94	0.89	
positive	0.88	0.82	0.72	

****PM: Partial Mastectomy, TM: Total Mastectomy**

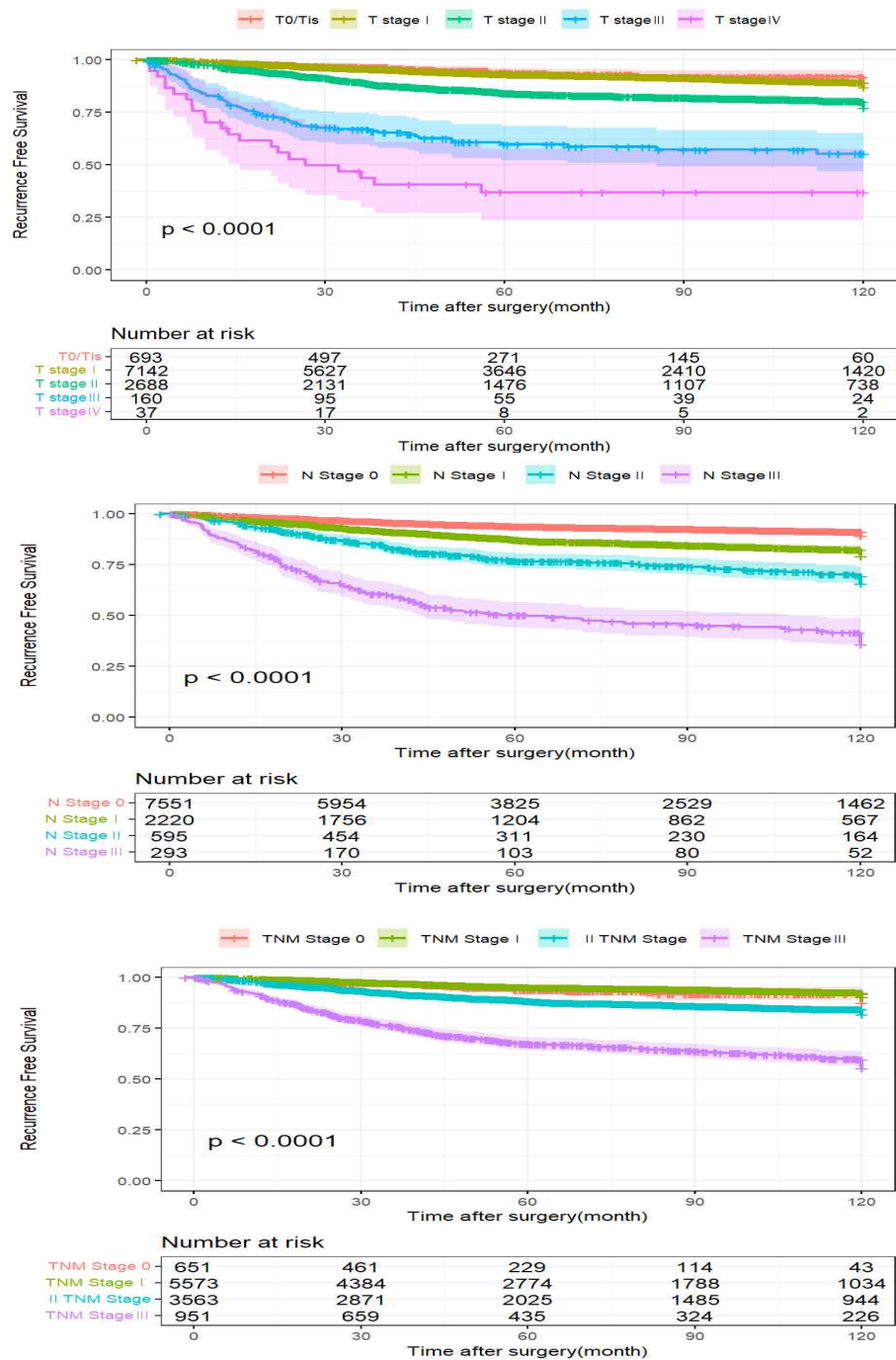


Figure 3.Kaplan-Meier Curve & Log rank Test

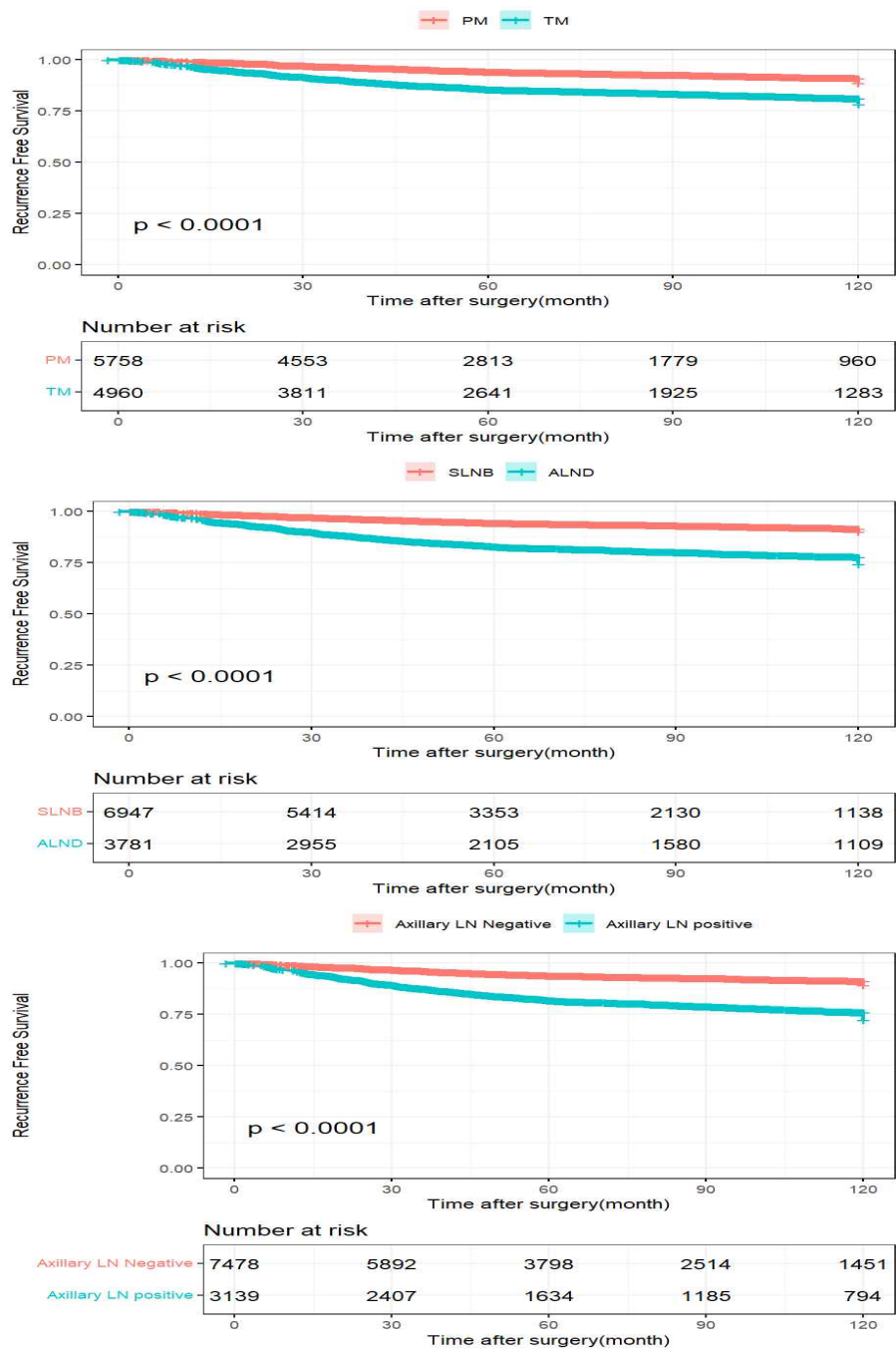


Figure 3.Kaplan-Meier Curve & Log rank Test

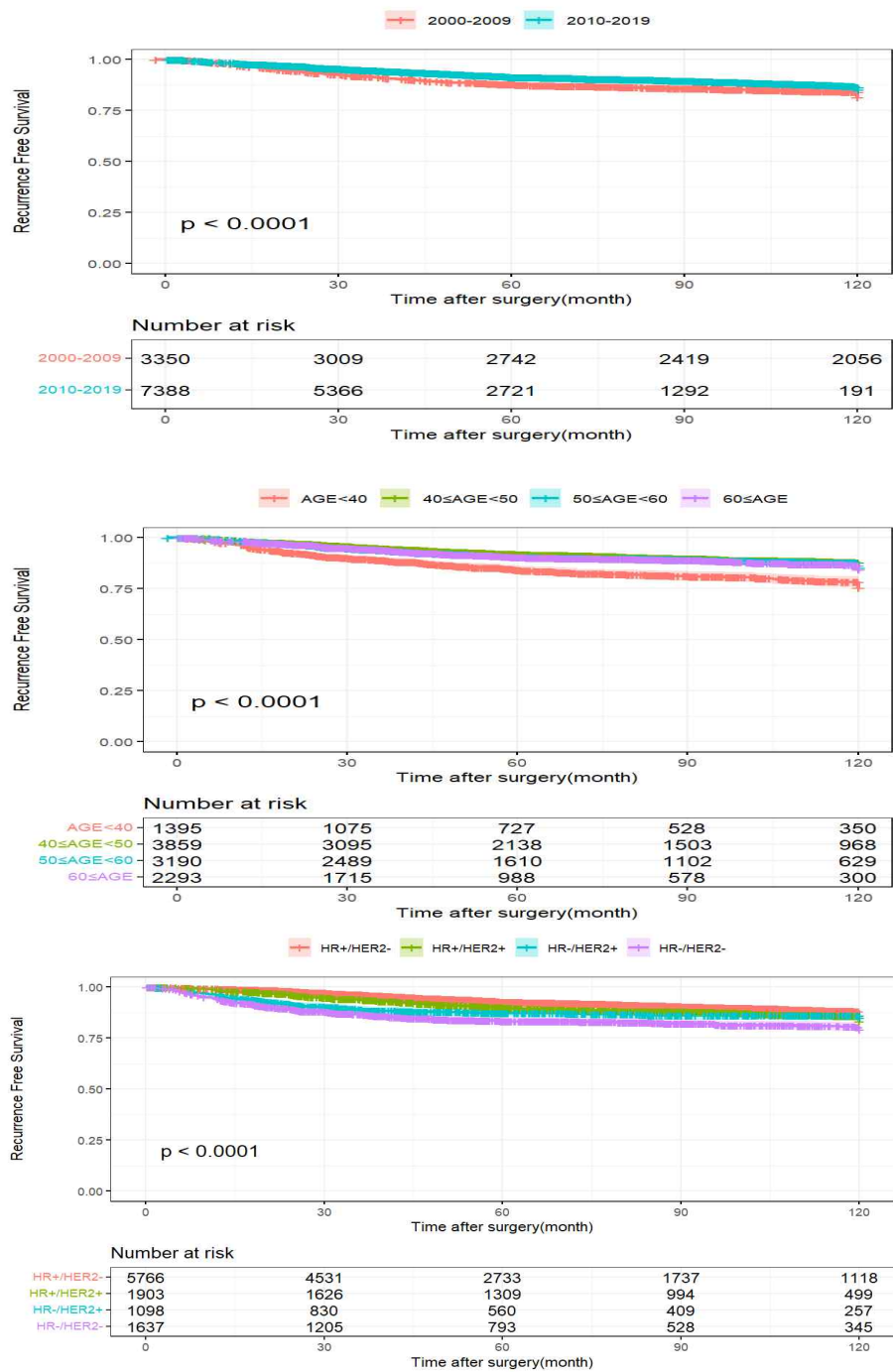


Figure 3. Kaplan-Meier Curve & Log rank Test

2.2. 유방암 수술 후 암 재발 상대위험도

콕스 비례 위험 모델을 이용하여 유방암 수술 후 재발 상대위험도를 확인하였다. 다변수 콕스 회귀모델은 단변량 cox 회귀분석에서 통계적으로 유의미한 변수들 중 다중공선성을 확인하여 VIF 10 이하를 만족하는 변수들을 넣어 분석하였다.

단변량 콕스 회귀모델에서 통계적으로 유의미한 변수는 다음과 같다. 유방암 수술 후 암 재발 상대위험도는 2000-2009년도 환자 대비 2010-2019년도 환자가 0.71배(95% CI 0.63-0.80) 낮았으며, BMI 정상 대비 과체중일 때 1.18배(95% CI 1.04-1.35) 비만일 때 1.12배(95% CI 0.86-1.46) 높았다. 연령이 30대 이하인 군 대비 40대 0.51배(95% CI 0.43-0.60) 50대 0.55(95% CI 0.46-0.65) 60대 이상인 경우 0.57배(95% CI 0.47-0.69)로 연령이 증가할수록 암 재발 상대위험도가 낮아졌다. 유방암 분자 아형에서 HR+/HER2-군 대비 HR+/HER2+군은 1.35(95% CI 1.15-1.58)배 높았고 HR-/HER2+군은 1.57(95% CI 1.30-1.90)배 높았으며, HR2-HER2-군은 2.17배(95% CI 1.87-2.53)로 암 재발 상대위험도가 가장 높았다. T stage에서 0기, Tis 대비 2기에서 2.56(95% CI 1.83-3.58)배 3기에서 8.02(95% CI 5.34-12.1)배, 4기에서 16.4(95% CI 9.66-27.8)배로 병기가 진행될수록 암 재발 상대위험도가 증가했다. N stage에서 0기 대비 1기에서 2.14(95% CI 1.85-2.46)배 낮으며 2기에서 3.97(95% CI 3.31-4.77)배 3기에서 10.7(95% CI 8.9-12.8)배 높다. TNM stage에서 0기 대비 암 재발 상대위험도는 2기에서 1.98(95% CI 1.39-2.82) 3기에서 6.17(4.31-8.83)배 높았다. 내분비 치료를 받지 않은 군 대비 받은 군에서 0.60(95% CI 0.53-0.68)배 암 재발 위험도가 낮았다. 수술 후 항암치료를 받지 않은 군 대비 받은 군에서 1.46(95% CI 1.29-1.66)배 높았다. 유

방암 부분 절제술을 받은 환자군 대비 전절제술을 받은 환자는 2.41(95% CI 2.12-2.73)배 높았다. 감시 림프절 생검을 받은 환자군 대비 액와림프절 절제술을 받은 환자의 3.13(95% CI 2.77-3.54)배 높았다. 액와 림프절 음성인 환자군 대비 양성인 환자의 암 재발 상대위험도는 3.10(95% CI 2.75-3.49)배 높다. 모유 수유 여부, 출산 여부, 가족력, 방사선 치료 유무는 통계적으로 유의하지 않았다.

다변수 콕스 회귀모델에서 나머지 변수를 통제하였을 때 통계적으로 유의한 수술 후 암 재발 상대위험도의 결과는 다음과 같다. BIM정상 대비 과체중에서 1.18(95% CI 1.03-1.35)배 높았다. 30대 이하 대비 40대 0.6(95% CI 0.50-0.71)배, 50대 0.63(95% CI : 0.52-0.75)배, 60대 0.63(95% CI : 0.47-0.70)배 낮았다. 유방암 분자 아형에서 HR+/HER2-군 대비 HR+/HER2+군은 1.21(95% CI : 1.02-1.43)배 높았고, HR+/HER2-군은 1.52(95% CI : 1.08-2.14)배 높았으며, HR2-HER2-군은 2.93(95% CI : 2.13-4.04)배로 가장 높았다. T stage에서 0기, Tis군 대비 2기에서 2.06(95% CI : 1.42-2.98)배, 3기에서 4.02(95% CI : 2.57-6.28)배 높았으며 4기에서 5.52(95% CI : 3.07-9.91)배로 가장 높았다. N stage에서 0기 대비 1기에서 1.6(95% CI : 1.30-1.97)배, 2기에서 2.83(95% CI : 2.17-3.65)배 이며, 3기에서 6.86(95% CI : 5.25-9.02)배로 가장 높았다. 방사선 치료를 받지 않은 군 대비 방사선 치료를 받은 군에서 0.81(95% CI : 0.68-0.98)배 낮았다. 수술 후 항암치료를 받지 않은 군 대비 항암치료를 받은 군에서 0.71(95% CI : 0.61-0.83)배 낮았다. 유방 부분절제술을 받은 환자군 대비 유방 전절제술을 받은 환자는 1.41(95% CI : 1.17-1.68)배 높았다. 감시림프절 생검을 받은 환자군 대비 액와 림프절 광청술을 받은 환자가 1.56(95% CI : 1.28-1.9)배 높았다.

Table4. Hazard ratio for Recurrence of Breast cancer.

Variables	Univariate analysis		Multivariable analysis	
	Hazard Ratio(95% CI)	P-value	Hazard Ratio(95% CI)	P-value
Year of operation				
2000-2009	1		1	
2010-2019	0.71(0.63-0.80)	<0.001	1.1(0.95-1.28)	0.20
BMI				
<25	1		1	
<30	1.18(1.04-1.35)	0.01	1.18(1.03-1.35)	0.016
≤30	1.12(0.86-1.46)	0.40	1.06(0.8-1.39)	0.70
Age				
<40	1		1	
40-49	0.51(0.43-0.60)	<0.001	0.6(0.5-0.71)	<0.001
50-59	0.55(0.46-0.65)	<0.001	0.63(0.52-0.75)	<0.001
≥60	0.57(0.47-0.69)	<0.001	0.57(0.47-0.70)	<0.001
Breast feeding				
no	1			
yes	1.05(0.92-1.19)	0.50		
Parity				
no	1			
yes	0.86(0.68-1.1)	0.20		
Presence of family history of cancer				
no	1			
yes	0.89(0.79-1.01)	0.065		
Subtype				
HR+/HER2-	1		1	
HR+/HER2+	1.35(1.15-1.58)	<0.001	1.21(1.02-1.43)	0.028
HR+/HER2-	1.57(1.30-1.90)	<0.001	1.52(1.08-2.14)	0.016
HR2-HER2-	2.17(1.87-2.53)	<0.001	2.93(2.13-4.04)	<0.001
T stage				
T0/Tis	1		1	
I	1.19(0.86-1.67)	0.30	1.42(0.99-2.03)	0.058
II	2.56(1.83-3.58)	<0.001	2.06(1.42-2.98)	<0.001
III	8.02(5.34-12.1)	<0.001	4.02(2.57-6.28)	<0.001
IV	16.4(9.66-27.8)	<0.001	5.52(3.07-9.91)	<0.001

Table4. Hazard ratio for Recurrence of Breast cancer(continued).

Variables	Univariate analysis		Multivariable analysis	
	Hazard Ratio(95% CI)	P-value	Hazard Ratio(95% CI)	P-value
N stage				
0	1		1	
I	2.14(1.85-2.46)	<0.001	1.61(1.3-1.97)	<0.001
II	3.97(3.31-4.77)	<0.001	2.83(2.17-3.65)	<0.001
III	10.7(8.9-12.8)	<0.001	6.86(5.25-9.02)	<0.001
TNM Stage				
0	1			
I	0.83(0.58-1.19)	0.3		
II	1.98(1.39-2.82)	<0.001		
III	6.17(4.31-8.83)	<0.001		
Endotherapy				
None	1		1	
Done	0.6(0.53-0.68)	<0.001	1.1(0.81-1.49)	0.60
Radiation therapy				
None	1		1	
Done	1.01(0.89-1.14)	>0.9	0.81(0.68-0.98)	0.03
Postoperative				
Chemotherapy				
None	1		1	
Done	1.46(1.29-1.66)	<0.001	0.71(0.61-0.83)	<0.001
Breast operation				
PM	1		1	
TM	2.41(2.12-2.73)	<0.001	1.41(1.17-1.68)	<0.001
Axillary operation				
SLNB	1		1	
ALND	3.13(2.77-3.54)	<0.001	1.55(1.28-1.89)	<0.001
Axillary Lymph node				
negative	1			
positive	3.1(2.75-3.49)	<0.001		

****PM: Partial Mastectomy, TM: Total Mastectomy**

3. 유방암 분자 아형에 따른 암 재발 위험도 하위그룹 분석

치료변수를 제외한 변수들로 유방암 분자 아형에 따른 암 재발 상대위험도 하위그룹 분석을 시행하였다.

HR+/HER2-군 대비 HR-/HER2+군의 암 재발 위험도는 2010년 이전에 수술 받은 경우(1.22, 95% CI : 0.92-1.62)보다 2010년 이후에 수술 받은 경우(1.81, 95% CI : 1.38-2.38)가 컸다. HR+/HER2-군 대비 HR-/HER2-군의 암 재발 위험도는 2010년 이전에 수술 받은 경우(1.57, 95% CI : 1.25-1.98)보다 2010년 이후에 수술 받은 경우(2.53, 95% CI : 2.04-3.14)가 컸다.

HR+/HER2-군 대비 HR+/HER2+군의 암 재발 위험도는 N stage 0기 1.15(95% CI : 0.90-1.47), 1기 1.13(95% CI : 0.82-1.55), 2기 1.66(95% CI : 1.07-2.57), 3기 0.55(95% CI : 0.32-0.96) 이며 이는 통계적으로 유의하다($p=0.018$). HR+/HER2-군 대비 HR-/HER2+군의 암 재발 위험도는 N stage 0기 1.73(95% CI : 1.32-2.27), 1기 1.26(95% CI : 0.85-1.86), 2기 1.64(95% CI : 0.96-2.81), 3기 0.59(95% CI : 0.29-1.21) 이며 이는 통계적으로 유의하다($p=0.022$). HR+/HER2-군 대비 HR-/HER2-군의 암 재발 위험도는 N stage 0기 1.88(95% CI : 1.51-2.34), 1기 2.97(95% CI : 2.21-4.01), 2기 2.26(95% CI : 1.32-3.87), 3기 4.87(95% CI : 3.07-7.73) 이며 이는 통계적으로 유의하다($p=0.002$).

BMI, 연령, 모유 수유 경험, 출산 경험, 암 가족력, T stage, TNM stage, 액와 림프절 전이 여부에 따라 유방암 분자 아형이 재발에 미치는 영향은 통계적으로 유의미하게 다르지 않다.

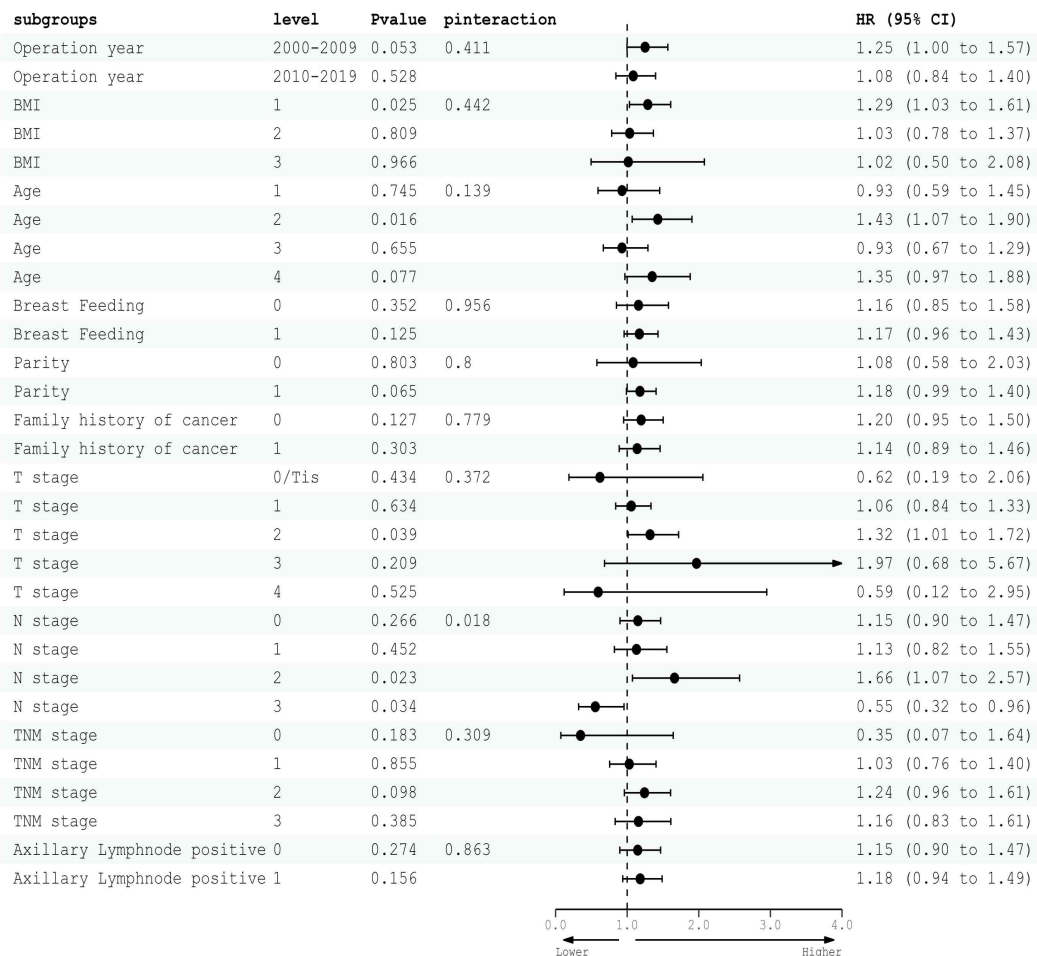


Figure.4 HR+/HER2-대비 HR+/HER2+ HR forest plot

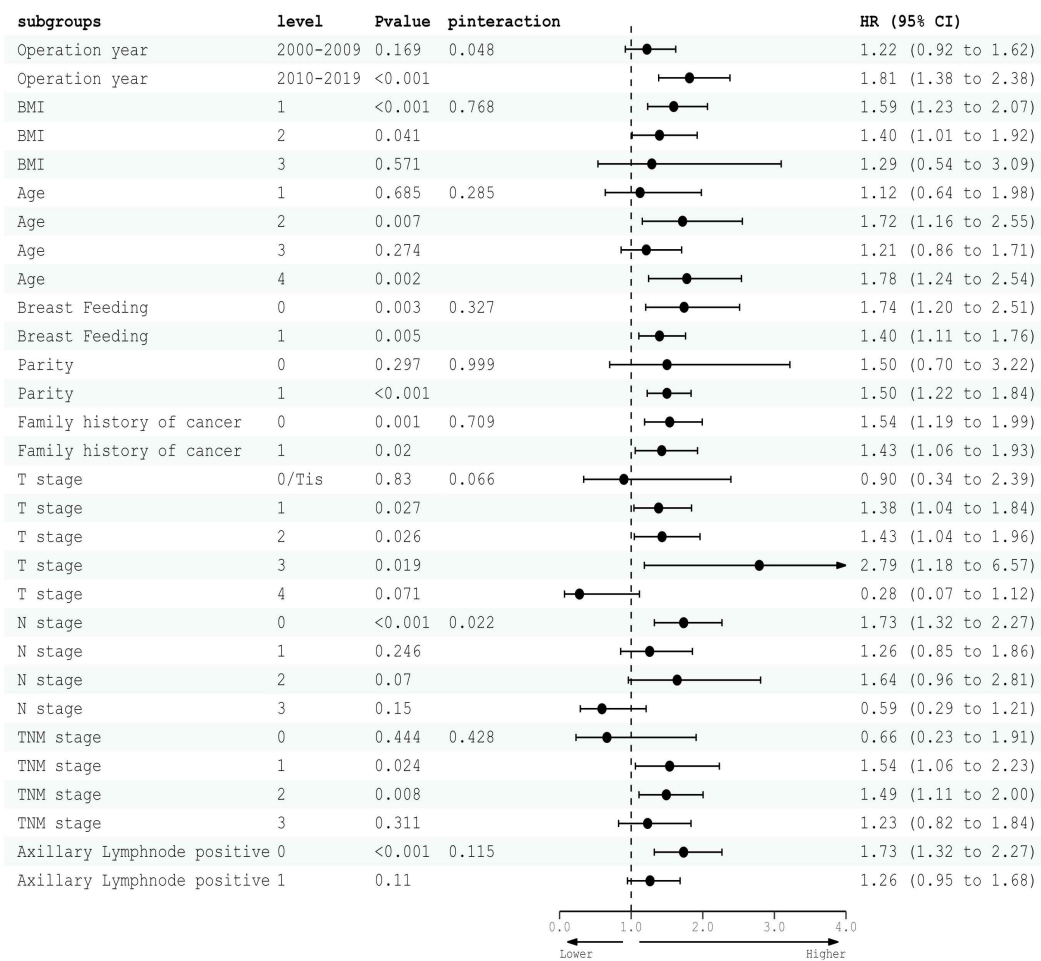


Figure.5 HR+/HER2-대비 HR-/HER2+ HR forest plot

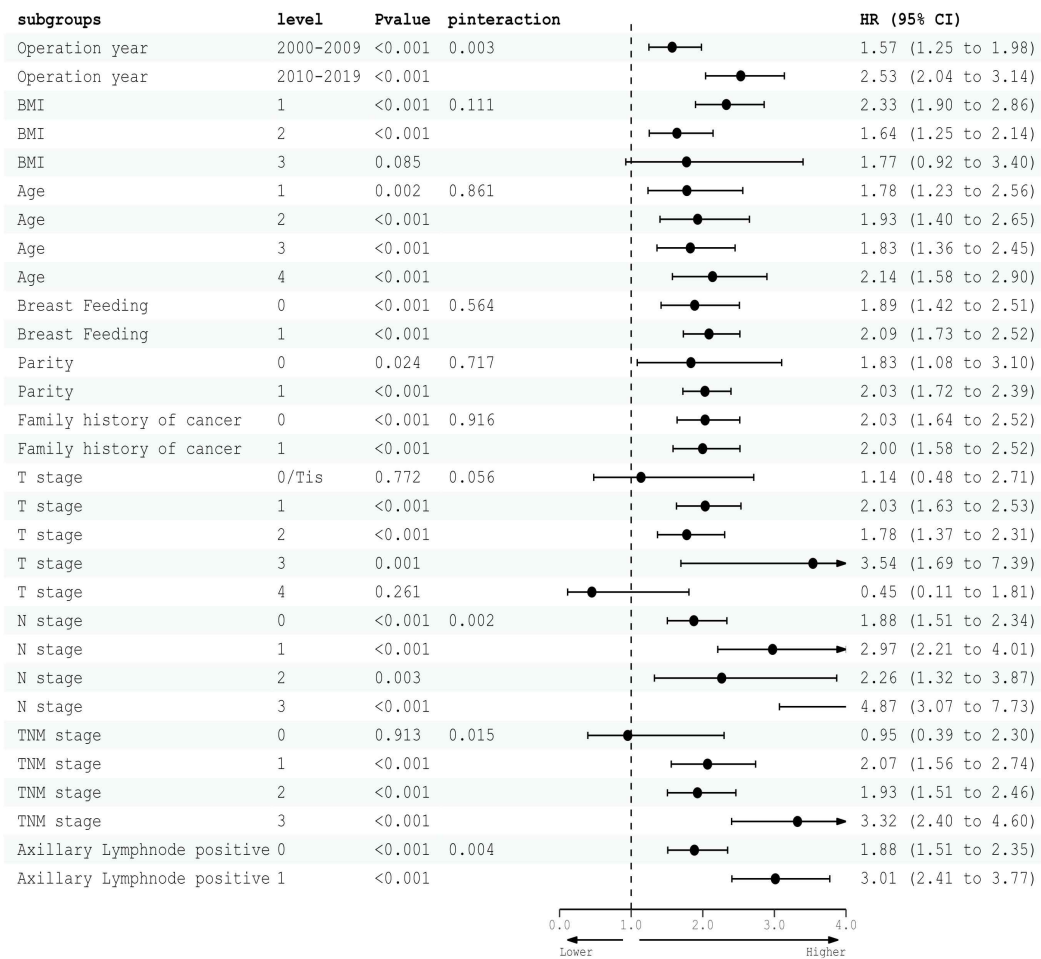


Figure.6 HR+/HER2- 대비 HR-/HER2- HR forest plot

4. Cox Proportional Hazards Model, Random Survival forest, XGBSE를 이용한 재발예측 모델 비교

Cox Proportional Hazards Model, Random Survival forest와 XGBSE를 이용하여 암 재발예측 모델을 만들고 C-index로 예측력을 비교하였다. Random Survival forest와 XGBSE 모델의 변수중요도를 확인하였다.

Cox Proportional Hazards Model의 train set에서 C-index 값은 0.716이고 test set에서 C-index 값은 0.728이다.

Random Survival forest를 이용한 모델의 하이퍼파라미터는 5-fold 그리드 탐색으로 최적의 파라미터 'max_features': 5, 'min_samples_split': 2, 'n_estimators': 100로 설정하였다. train set에서 C-index 값은 0.886이고 test set에서 C-index 값은 0.721이다. 변수중요도는 연령, BMI, TNM stage, T stage, 수술 후 항암치료 여부, 가족력, 모유 수유 여부, 유방암 수술 종류, 방사선 치료 여부, N stage, 유방암 분자 아형 종류 순으로 중요하게 나타난다.

XGBSE를 이용한 모델의 하이퍼파라미터는 5-fold 그리드 탐색으로 'aft_loss_distribution': 'normal', 'aft_loss_distribution_scale': 1.5, 'colsample_bynode': 0.6, 'learning_rate': 0.1, 'max_depth': 8, 'min_child_weight': 60, 'objective': 'survival:aft', 'subsample': 0.5, 으로 설정하였다. train set에서 C-index 값은 0.769이고 test set에서 C-index 값은 0.692이다. 변수중요도 순서는 연령, 가족력, 유방암 분자 아형 종류, 유방암 수술 종류, BMI, 수술 시행 여부, TNM stage, 방사선 치료 여부, 수술 후 항암치료 여부, T stage 순으로 중요하게 나타난다.

Table 5. C-index of RSF Model, XGBSE Model.

	C-index in Train set	C-index in Test set
Cox PH	0.716	0.728
Random Survival Forest	0.886	0.721
XGBSE	0.769	0.692

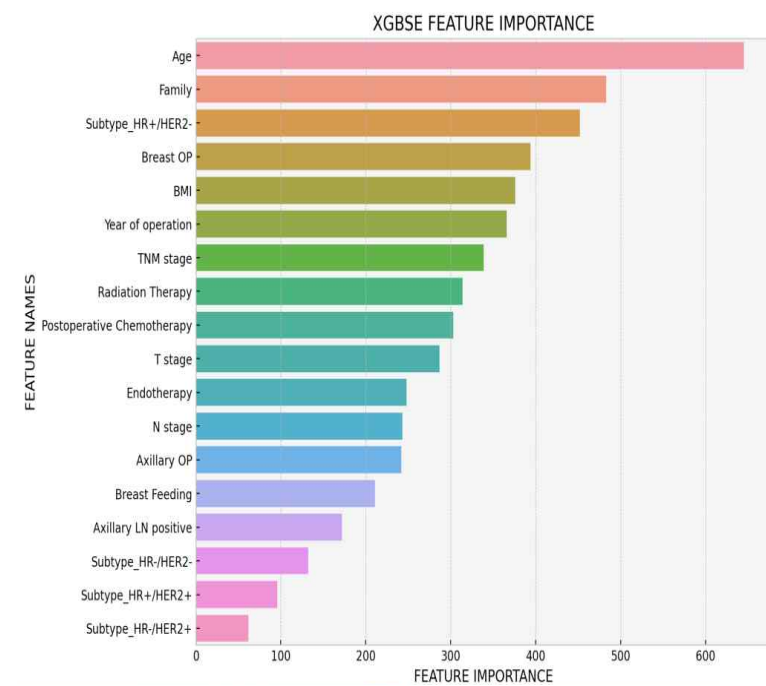
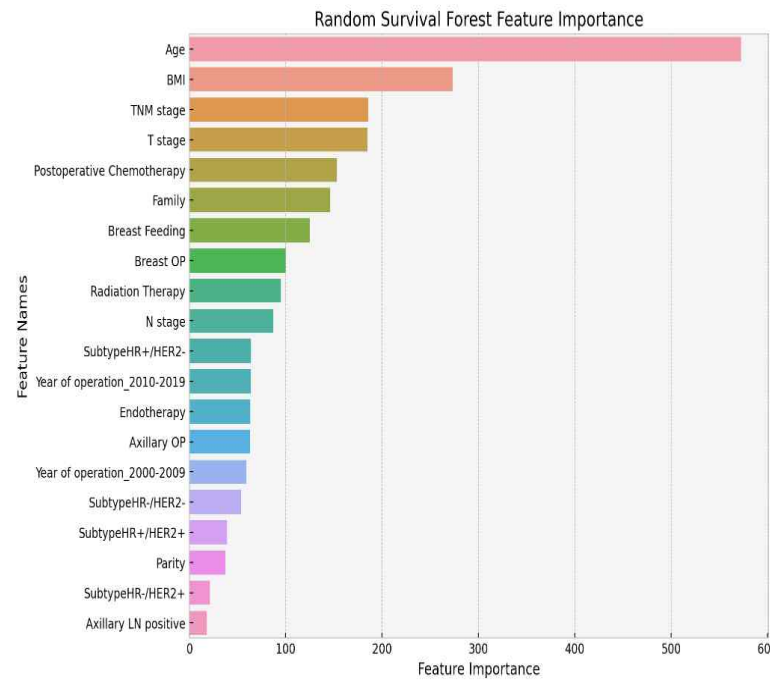


figure.7 Feature Importance plot of RSF, XGBSE Model.

V. 고찰

1. 연구 결과에 대한 고찰

2010년 이후에 수술받은 환자들의 무 재발 생존율이 2010년 이전에 수술받은 환자들보다 높게 나타났으며($p < 0.001$) 이는 조기 진단의 증가와 수술 및 치료 방법의 발전(한국 유방암학회, 2020)에 따른 결과로 사료 된다. 30대 이하 환자들이 가장 낮은 무 재발 생존율을 보였고 이는 유방절제술 후 10년 국소 재발률 및 전신 재발률이 35세 이하인 군에서 높았다는 선행 연구와 일치한다(Kim et al., 2001). 유방암 분자 아형은 HR-/HER2-, HR-/HER2+, HR+/HER2+, HR+/HER2- 순으로 무 재발 생존율이 낮았으며, 삼중 음성인 경우의 3년(0.87), 5년(0.83), 10년(0.79)의 무 재발 생존율이 가장 낮았다($p < 0.001$). 이는 삼중음성 유방암의 예후가 가장 좋지 않은 선행 연구와 일치한다(Howlader N, 2018). T stage, N stage, TNM stage 모두 병기가 진행될수록 무 재발 생존율이 감소했다($p < 0.001$).

내분비 치료를 받은 경우, 수술 후 항암치료를 받지 않은 경우는 그렇지 않은 경우보다 무 재발 생존율이 높았다($p < 0.001$). 이는 호르몬 수용체 양성 유방암에 내분비 치료를 시행하며 호르몬 수용체 양성유방암의 예후가 호르몬 수용체 음성 유방암보다 좋기 때문(Howlader N, 2018)으로 생각된다. 유방 전 절제술을 받은 환자는 유방 부분 절제술을 받은 환자보다 무 재발 생존율이 낮았다(< 0.001). 액와 림프절 광청술을 받은 환자는 액와 감시 림프절 생검술을 받은 환자보다 무재발 생존율이 낮았다(< 0.001). 액와 림프절 양성인 환자의 무 재발 생존율은 액와 림프절 음정보다 낮다(< 0.001). 이는 액와 림프절 양성으로 이미 액와 림프절에 전이가 있는 경우 액와 림프절 광청술

을 시행하기 때문에 액와 림프절 광청술을 받은 환자의 예후가 액와 감시 림프절 절제술을 받은 환자보다 좋지 않기 때문이다.

다변수 콕스 회귀모델에서 나머지 변수를 통제하였을 때 암 재발 위험도가 BMI 정상 대비 과체중에서 1.18(95% CI : 1.03-1.35)배 높았다($p=0.016$). 연령에서는 30대 이하 대비 연령이 증가할수록 암재발 위험도가 낮아졌다($p<0.001$). 유방암 분자 아형에서 HR+/HER2-군 대비 HR+/HER2+군은 1.21(95% CI : 1.02-1.43)배 높았고($p=0.028$), HR+/HER2-군은 1.52(95% CI : 1.08-2.14)배 높으며($p=0.016$), HR-/HER2-군은 2.93(95% CI : 2.13-4.04)배로 가장 높았다($p<0.001$). T stage, N stage 모두 병기가 진행될수록 암 재발 위험도가 증가한다($p<0.001$). T stage 4기에서 5.52(95% CI : 3.07-9.91)배로 가장 높았다($p<0.001$). N stage 3기에서 6.86(95% CI : 5.25-9.02)배로 가장 높았다($p<0.001$). 방사선 치료를 받지 않은 군 대비 방사선 치료를 받은 군에서 0.81(95% CI : 0.68-0.98)배 낮았다($p<0.001$). 수술 후 항암치료를 받지 않은 군 대비 항암치료를 받은 군에서 0.71(95% CI : 0.61-0.83)배 낮았다($p<0.001$).

이는 진단 시 연령이 낮을수록, 유방암 분자 아형이 HR+/HER2-대비 HR+/HER2+, HR-/HER2+, HR-/HER2-일수록, T stage, N stage가 진행될수록, 방사선 치료와 항암치료를 받지 않을수록 재발 위험도가 높은 선행 연구의 결과와 일치한다(Cheun et al., 2023).

유방 부분절제술을 받은 환자군 대비 유방 전절제술을 받은 환자는 1.41(95% CI : 1.17-1.68)배 높았다($p<0.001$). 감시 림프절 생검을 받은 환자군 대비 액와 림프절 광청술을 받은 환자가 1.56(95% CI : 1.28-1.9)배 높았다($p<0.001$). 암의 크기가 크고 병기가 진행되어있는 경우 유방암 전 절제술을 시행하기 때문에(Paik, 2014) 위와 같은 결과가 나왔으리라 생각된다.

유방암 분자 아형에 따른 암 재발 상대위험도 하위그룹 분석 결과 수술받은 연도와 N stage에 따라 유방암 분자 아형이 암 재발에 미치는 상대위험이 차이가 있다. 2010년 이후에 수술받은 경우가 2010년 이전에 수술받은 경우보다 유방암 분자 아형이 암 재발에 미치는 영향이 크다. N stage 2기, 3기인 경우에서 유방암 분자 아형이 암 재발에 미치는 영향이 크다.

머신러닝 모델에서 데이터 세트에서 샘플을 추출할 때 무작위로 추출하는 과정에서 sampling variation이 발생하며 이는 모델 평가의 불확실성으로 이어진다. 이를 완화하기 위해 교차검증을 시행하였으며 머신러닝 분석 결과 Cox 비례위험 모델보다 머신러닝 모델의 예측력이 낮았다. RSF 모델의 성능은 train set(C-index 0.886)이 Test set(C-index 0.721)보다 좋으며 이는 RSF 모델이 Train set에 과적합 되었다고 볼 수 있다. XGBSE 모델은 Train set(C-index 0.769)과 Test set(C-index 0.692) 모두에서 비슷한 성능을 보이지만 예측력이 가장 낮았다.

변수중요도의 순서는 두 모델에 차이가 있었는데, 다변수 콕스 회귀모델에서 암 재발 위험도에 유의한 변수들인 연령, 유방암 분자 아형, T stage, N stage, 방사선 치료 여부, 항암치료 여부가 RSF모델의 변수중요도 순서의 상위권에 더 많이 위치한다.

2. 연구의 강점과 제한점

본 연구의 강점은 다음과 같다. 첫째, 유방암 분자 아형과 병기에 따른 유방암 재발 위험도 차이가 있음을 확인하였고 이는 유방암 분자 아형과 병기의 정확한 진단이 유방암 재발 위험을 낮추는데 중요함을 재고한다. 둘째, 본 연구는 국내 단일기관 자료를 이용해서 유방암 재발 예측 머신 러닝 분석을 진행했다는 것에 의의가 있다. 셋째, 기존에 밝혀진 유방암 재발 위험 요인을 확인하였다. 진단 시 연령, 유방암 분자 아형, TNM stage, 내분비 치료 여부, 방사선 치료 여부, 수술 후 항암치료 여부, 유방절제술 종류, 액와 림프절 절제술 종류, 액와 림프절 전이 여부가 유방암 재발에 유의한 변수임을 확인하였다.

한계점은 다음과 같다. 첫째, 외적 타당도 검증이 시행되지 않았다. 전체 유방암 환자에 대한 결과로 일반화하기에는 어려움이 있으므로 추후 다기관 자료를 이용한 연구가 필요할 것이다. 둘째, 높은 예측력의 모델을 구축하지 못했다. 이는 다중공선성이 높기 때문이다. 유방암 분자 아형과 병기의 근거하여 치료 방법을 선택하기 때문에 모델에 사용된 변수들간의 연관성이 높은 문제가 있었다. 추후 임상적 지표와 관련된 다른 변수들을 추가하여 모델을 구축해야한다. 셋째, 유방암 재발을 국소 재발, 구역 재발, 원격 재발을 모두 포함하여 정의였다. 따라서 각각의 부위별 재발 위험에 대한 차이는 확인할 수 없으므로 추후 연구가 필요할 것이다.

VI. 결론

본 연구는 유방암 수술 후 암 재발 여부에 따른 일반적 특성과 유방암 분자 아형에 따른 일반적 특성을 파악하고, 카플란 마이어 분석을 이용하여 유방암 수술 후 무 재발 생존율에 영향을 미치는 요인을 확인하였다. 콕스 비례 위험분석으로 유방암 재발 위험도를 확인하고 하위그룹 분석을 이용하여 분자 아형에 따른 유방암 재발 위험도가 변수 그룹 간에 유의미한 차이가 있는지 확인하였다. 머신러닝을 이용한 유방암 예측 모델을 랜덤 서바이벌 포레스트, XGBSE로 구축하고 예측력을 비교하였다.

Kaplan-Meier test 결과 수술 시행 연도가 2010년 이전일수록, 연령이 30대 이하인 경우, 유방암 분자 아형이 삼중 음성인 경우, N stage 병기가 진행될수록, 내분비 치료를 받지 않은 경우, 수술 후 항암치료를 받지 않은 경우, 유방 전 절제술을 받은 경우, 액와 림프절 광청술을 받은 경우, 액와 림프절 양성인 경우의 무 재발 생존율이 낮다.

콕스 비례위험분석 결과 다른 변수들을 통제하였을 때 진단 시 연령이 높을수록 재발 위험도가 감소하며, 유방암 분자 아형 HR+/HER2-군 대비 HR+/HER2+군, HR-/HER2+군, HR-/HER2-군일수록 재발 위험이 증가한다. T stage, N stage 병기가 진행될수록 재발 위험이 증가한다. 방사선 치료를 받는 경우, 수술 후 항암치료를 받은 경우는 그렇지 않은 경우에 비해 암 재발 위험이 낮다. 유방 부분 절제술을 받은 군 대비 유방 전절제를 받은 군의 암 재발 위험이 높으며, 액와 림프절 감시 생검술을 받은 군 대비 액와 림프절 광청술을 받은 군의 재발 위험이 높다.

하위그룹 분석 결과 수술 받은 연도와 N stage 병기에 따라 유방암 분자 아형에 따른 암 재발 위험도가 차이가 있음을 확인하였다. 특히 N stage 3기에서 유방암 분자 아형이 암 재발 위험도에 미치는 영향이 컸다.

머신러닝 분석 결과 RSF 모델이 XGBSE모델 보다 예측력이 컸으나 Cox 비례위험 모델보다는 예측력이 낮았다.

유방암 재발 위험 요인을 분류하고 유방암 재발을 예측하면 환자의 유방암 재발을 조기에 중재하고 유방암 재발율을 낮출 수 있을 것이다. 연구 결과를 기반으로 향후 대규모의 다기관 자료를 이용한 유방암 재발 예측 모델에 대한 연구가 필요하다고 생각된다.

참고문헌

한국유방암학회, 유방암 백서, 2020.

Tran, K.A., Kondrashova, O., Bradley, A. et al. Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med* 13, 152 (2021).

Cheun JH, Kim HK, Mood HG, Locoregional Recurrence Patterns in Patients With Different Molecular Subtypes of Breast Cancer, *JAMA Surg.* 2023;158(8):841-852.

Bevers TB, Helvie M, et al.. Breast Cancer Screening and Diagnosis, Version 3.2018, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw.* 2018 Nov;16(11):1362-1389.

백남선, 유방암 발생과 치료의 최근 동향, *Ewha Med J* 2014; 37(2):69-74 pISSN: 2234-3180, eISSN: 2234-2591

우주현, 임우성. 유방암의 내분비 요법. *Ewha Med J* 2014;37(2):83-91.

백남선, 유방암 발생과 치료의 최근 동향, *Ewha Med J* 2014; 37(2):69-74 pISSN: 2234-3180, eISSN: 2234-2591.

정승필, 우숙영, 이세경. 유방암 수술 후 호르몬 수용체와 나이에 따른 재발 위험률에 대한 연구. J Breast Dis 2013 November; 1(1): 15-20.

우주현, 임우성. 유방암의 내분비 요법. Ewha Med J 2014;37(2):83-91.

Zuo D, Yang L, Jin Y, Qi H, Liu Y, Ren L. Machine learning-based models for the prediction of breast cancer recurrence risk. BMC Med Inform Decis Mak. 2023 Nov 29;23(1):276.

Lafourcade, A., His, M., Baglietto, L. et al. Factors associated with breast cancer recurrences or mortality and dynamic prediction of death using history of cancer recurrences: the French E3N cohort. BMC Cancer 18, 171 (2018).

Voogd AC, Nielsen M, Peterse JL, et al. Differences in risk factors for local and distant recurrence after breast-conserving therapy or mastectomy for stage I and II breast cancer: pooled results of two large European randomized trials. J Clin Oncol. 2001;19(6):1688-97.

Pedersen RN, Esen BÖ, Mellemkjær L, Christiansen P, Ejlersen B, Lash TL, Nørgaard M, Cronin-Fenton D. The Incidence of Breast Cancer Recurrence 10-32 Years After Primary Diagnosis. J Natl Cancer Inst. 2022 Mar 8;114(3):391-399.

Pan H, Gray R, Braybrooke J, et al. EBCTCG. 20-year risks of breast-cancer recurrence after stopping endocrine therapy at 5 years. N Engl J Med. 2017;

377(19):1836-1846.

Clift AK, Dodwell D, Lord S, Petrou S, Brady M, Collins GS, Hippisley-Cox J.
Development and internal-external validation of statistical and machine
learning models for breast cancer prognostication: cohort study. *BMJ*. 2023
May 10;381:e073800. doi: 10.1136/bmj-2022-073800. PMID: 37164379;
PMCID: PMC10170264.

Beenken SW, Urist MM, Zhang Y, Desmond R, Krontiras H, Medina H, et al.
Axillary lymph node status, but not tumor size, predicts locoregional
recurrence and overall survival after mastectomy for breast cancer. *Ann Surg*
2003;237:732-738; discussion 738-739.

Hemant Ishwaran, Udaya B. Kogalur, Eugene H. Blackstone and Michael S. Lauer.
Random survival forests, *The Annals of Applied Statistics* 2008, Vol. 2, No.
3, 841-860 DOI: 10.1214/08-AOAS169 In the Public Domain

Lyman GH, Giuliano AE, Somerfield MR, Benson AB, 3rd, Bodurka DC, Burstein
HJ, et al. American Society of Clinical Oncology guideline recommendations
for sentinel lymph node biopsy in early-stage breast cancer. *J Clin Oncol*
2005;23:7703-7720.

Tianqi C, Carlos G. KDD '16: Proceedings of the 22nd ACM SIGKDD International
Conference on Knowledge Discovery and Data Mining August 2016 Pages 785-794.
Barke LD, Freivogel ME. Breast Cancer Risk Assessment Models and High-Risk

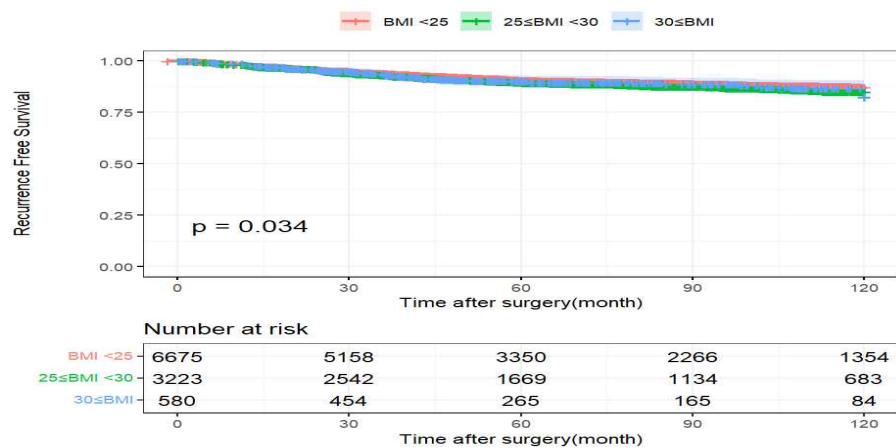
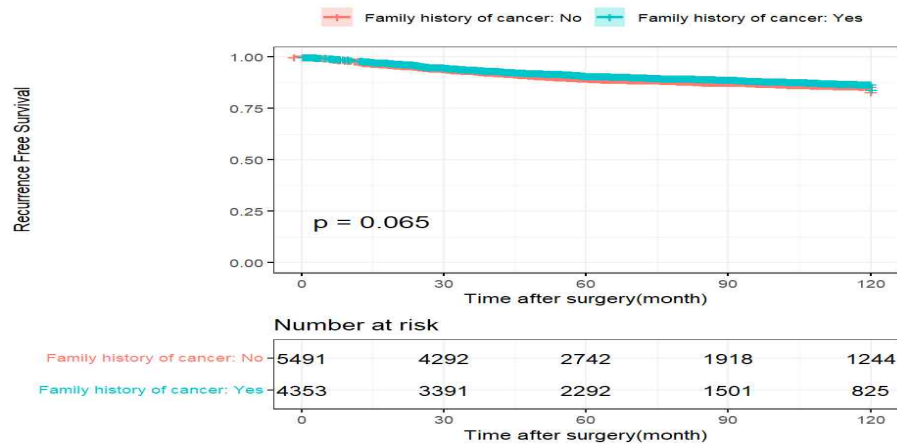
Screening. Radiol Clin North Am 2017;55:457-74.

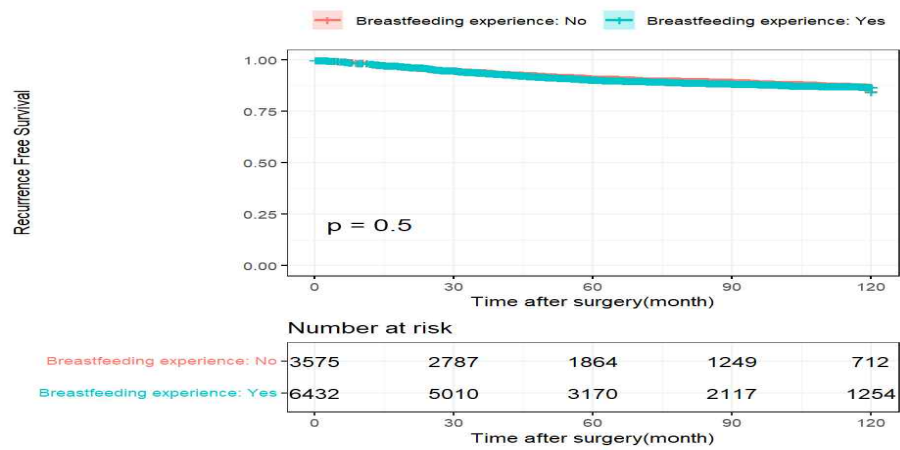
Partin JF, Mamounas EP. Impact of the 21-gene recurrence score assay compared with standard clinicopathologic guidelines in adjuvant therapy selection for node-negative, estrogen receptor-positive breast cancer. Ann Surg Oncol 2011;18:3399-3406.

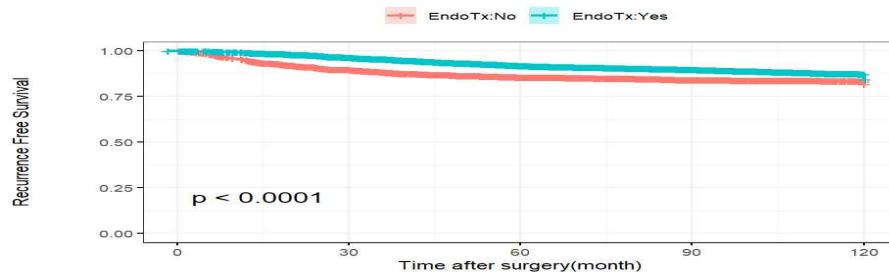
Steck, Harald, et al. On ranking in survival analysis: Bounds on the concordance index. Advances in neural information processing systems. 2008.

부록

1. 유방암 재발 카플란 마이어 커브 그래프



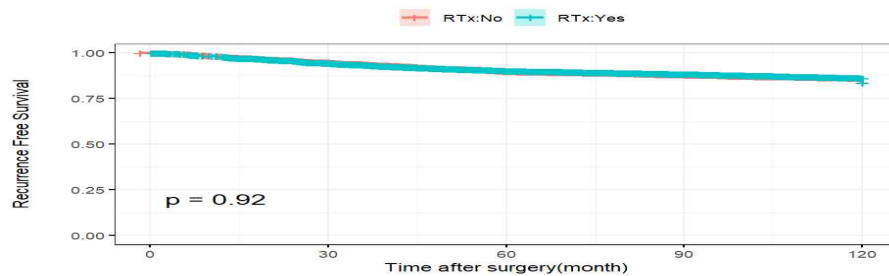




Number at risk

EndoTx:No	2948	2187	1386	958	619
EndoTx:Yes	7786	6185	4076	2752	1628
	0	30	60	90	120

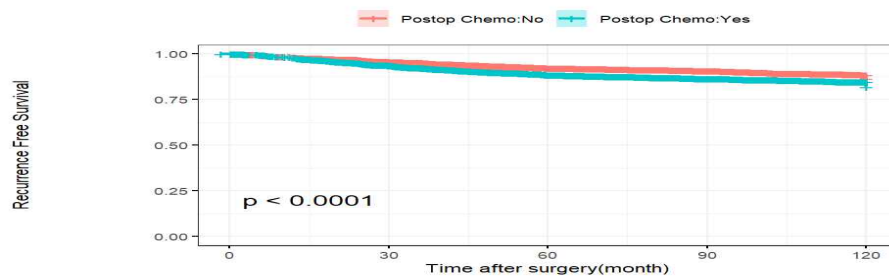
Time after surgery(month)



Number at risk

RTx:No	3525	2727	1833	1359	928
RTx:Yes	7206	5642	3626	2350	1319
	0	30	60	90	120

Time after surgery(month)



Number at risk

Postop Chemo:No	5060	3684	1992	1138	566
Postop Chemo:Yes	5670	4683	3467	2571	1680
	0	30	60	90	120

Time after surgery(month)

2. Best Parameter of RSF Model, XGBSE Model.

Best Parameter	
Random Survival Forest	'max_features': 5,
	'min_samples_split': 2,
	'n_estimators': 100
XGBSE	'aft_loss_distribution': 'normal',
	'aft_loss_distribution_scale': 1.5,
	'colsample_bynode': 0.6,
	'learning_rate': 0.1,
	'max_depth': 8,
	'min_child_weight': 60,
	'objective': 'survival:aft',
	'subsample': 0.5,

= ABSTRACT =

Risk Factors for Breast Cancer Recurrence and Recurrence Prediction Models in Patients with Breast Cancer Surgery

Haerim Lee

Department of Biostatistics

Graduate School of Public Health

Yonsei University

(Directed by Professor Chung Mo Nam, Ph.D.)

Despite the increased breast cancer survival rates due to early detection and improved medical technology, approximately 8-10% of breast cancer patients experience local recurrence, and 15-30% experience distant metastasis. Accurate prediction of breast cancer recurrence is crucial for enabling early intervention and improving overall survival rates. While research on breast cancer recurrence prediction using machine learning is actively underway, there is still a shortage of studies based on domestic data.

This study utilized retrospective data collected from 13,968 patients who were diagnosed with breast cancer and underwent breast surgery at Severance Hospital from January 1, 2000, to May 31, 2019, excluding

patients with pre-existing distant metastasis (120 patients) and those diagnosed with non-invasive breast cancer (2,523 patients). General characteristics based on breast cancer molecular subtypes and recurrence status were examined using the Chi-square test. Kaplan-Meier survival analysis was employed to analyze the recurrence-free survival rate after breast cancer surgery based on risk factors and treatment variables, with significance tested using the log-rank test. Cox Proportional Hazard models were used to analyze the risk factors for cancer recurrence after breast cancer surgery. Subgroup analysis was conducted to determine if there were differences in cancer recurrence risk according to breast cancer molecular subtypes. Predictive models for breast cancer recurrence were constructed using Cox Proportional Hazard models, Random Survival Forest, and XGBSE, and their goodness of fit and variable importance were compared. Data analysis was performed using R 4.3.4 and Python 3.12.2.

The recurrence rate of breast cancer was approximately 11%, with 89% of patients remaining recurrence-free. HR+/HER2- subtype accounted for 54%, while HR+/HER2+ accounted for 18%, HR-/HER2+ for 10%, and HR-/HER2- for 15%. Recurrence-free survival rates showed significant differences according to the year of surgery, age at diagnosis, breast cancer molecular subtype, N stage, endocrine therapy, adjuvant chemotherapy, and surgical method. Notably, patients who underwent surgery after 2010 had higher recurrence-free survival rates than those who underwent surgery before 2010 ($p < 0.001$), and patients aged under 30 showed the lowest recurrence-free survival rates ($p < 0.001$). Among

breast cancer molecular subtypes, triple-negative patients showed the lowest recurrence-free survival rates, and as T stage, N stage, and TNM stage progressed, recurrence-free survival rates decreased ($p < 0.001$).

Cox proportional hazard analysis revealed that higher age at diagnosis was associated with decreased recurrence risk ($p < 0.001$). The risk of recurrence was observed to be higher in the order of HR+/HER2-, HR+/HER2+, HR-/HER2+, and HR-/HER2- ($p < 0.001$). Recurrence risk also increased with advancing T stage and N stage ($p < 0.001$). Patients who received radiation therapy or adjuvant chemotherapy had lower recurrence risk compared to those who did not ($p < 0.001$). Patients who underwent breast-conserving surgery had lower recurrence risk compared to those who underwent total mastectomy ($p < 0.001$), and those who underwent axillary lymph node dissection had higher recurrence risk compared to those who underwent sentinel lymph node biopsy ($p < 0.001$).

Subgroup analysis of relative risk of cancer recurrence according to breast cancer molecular subtypes revealed that surgery after 2010 had a greater impact on the effect of breast cancer molecular subtypes on cancer recurrence than surgery before 2010 ($p = 0.048$). The effect of breast cancer molecular subtypes on cancer recurrence was significant in N stage 2 and 3 cases ($p = 0.018$).

Machine learning analysis results showed that machine learning models had lower predictive power than Cox proportional hazard models. The performance of the Random Survival Forest model was better on the training set (C-index 0.886) than on the test set (C-index 0.721), indicating overfitting to the training set. Although the XGBSE model

showed similar performance in both the training set (C-index 0.769) and test set (C-index 0.692), its predictive power was the lowest.

This study's significance lies in confirming the risk of breast cancer recurrence according to breast cancer molecular subtypes and staging and developing a predictive model for breast cancer recurrence using domestic single-center data. It underscores the importance of establishing patient follow-up plans considering breast cancer molecular subtypes and staging. However, since only data from a single institution were used, further research using large-scale multi-center data is needed to validate the study results.

Key words: Cancer recurrence, Molecular subtype, Cox proportional hazards model, Machine Learning