



Risk Prediction Model for Lung Cancer Screening

폐암검진에서의 위험예측모델

Tae Jung Kim, MD^{1*} , Hyae Young Kim, MD²,
Jin Mo Goo, MD^{3,4}, Joo Sung Sun, MD⁵

¹Department of Radiology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

²Department of Diagnostic Radiology, National Cancer Center, Goyang, Korea

³Department of Radiology, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Korea

⁴Institute of Radiation Medicine, Seoul National University Medical Research Center, Seoul, Korea

⁵Department of Radiology, Ajou University College of Medicine, Suwon, Korea

Lung cancer screening in high-risk subjects using low-dose CT can reduce mortality by 20%. Current evidence suggests that the development of a risk prediction model for lung cancer is one of the major advances in lung cancer screening. Herein, we review the technical requirements for evaluating different risk prediction models. Moreover, we describe the major lung cancer risk prediction models reported, and the results of lung cancer screening using these models.

Index terms Risk Assessment; Lung Cancer; Screening; Computed Tomography, X-Ray

서론

저선량 CT를 이용한 폐암검진은 금연 이외에 폐암 사망률을 감소시킬 수 있는 검증된 유일한 방법이다. 2011년 미국의 폐암검진 National Lung Screening Trial (이하 NLST)은 55~74세의 30갑년 이상의 흡연력을 지닌 무증상 고위험군 흡연자 및 15년 이내의 금연자에서 저선량 CT가 흉부 X선과 비교하여 20%의 사망률 감소를 가져올 수 있다고 보고하였다(1). NLST 연구 결과는 US Preventive Services Task Force (이하 USPSTF)를 포함하여, American Association of Thoracic Surgery, American Cancer Society, National Comprehensive Cancer Network (이하 NCCN), American College of Chest Physicians, American Society of Clinical Oncology 등과 같은 NLST와 유사한 연령과 흡연력을 지닌 사람을 대상으로 하는 다수의 폐암검진 권고안의 근거가 되었다(2-5). NLST 기준은 임상시험을 위해서는 현실적이고 적용이 용이한 장점이 있으나, 연속변수를 이분화함으로써 정보의 손실이 발생하고 다수의 변수가 반영되지 못하며 비선형효과(nonlinear effect)가 고려되지 못하므로,

Received May 20, 2019

Revised July 23, 2019

Accepted July 25, 2019

*Corresponding author

Tae Jung Kim, MD

Department of Radiology,
Samsung Medical Center,
Sungkyunkwan University
School of Medicine,
81 Irwon-ro, Gangnam-gu,
Seoul 06351, Korea.

Tel 82-2-3410-0715

Fax 82-2-3410-2559

E-mail taejung.kim1@gmail.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ORCID iD

Tae Jung Kim

[https://](https://orcid.org/0000-0003-2429-3533)

orcid.org/0000-0003-2429-3533

인구집단 수준에서 검진 목적의 위험도 평가에는 제한점이 있다. 또한 NLST 결과에서 알 수 있듯이, 1명의 폐암 사망을 예방하기 위해서는 320명의 검진이 필요하고(1), 해마다 검진 대상자의 약 20%에서 추적검사가 필요하였으며 이중 폐암이 발견된 비율은 1%에 불과하였다(6). 이러한 결과는 현재의 NLST의 검진 대상 선택기준으로는 다수의 위양성이 발생할 수밖에 없고, 불필요한 침습적 검사를 유발하여 결과적으로는 잠재적인 위해를 가져올 수 있음을 의미한다.

위험예측모델(risk prediction model)은 특정 위험 인자를 가진 개인에게 일정 기간 내에 특정 결과(outcome)가 발생할 확률을 예측하기 위한 모델이다. 우수한 위험예측모델을 검진에 적용하면, 검진 대상자를 보다 정교하게 선별할 수 있고, 이들에게서 폐암 위험요소에 해당되는 생활습관(예: 흡연)을 즉각 교정할 수 있다. USPSTF에서는 검진의 이득(benefit)은 최대화하고 위해(harm)는 최소화하기 위해서 폐암 고위험군을 정확히 찾아내는 것이 매우 중요함을 강조하고 있다(2). 최근 폐암 발생 및 사망률을 예측하기 위한 다양한 위험예측모델들이 개발되고 있다. 이 원고에서는 위험예측모델을 평가하는데 필요한 기술적인 요건과 지금까지 보고된 주요 폐암 위험예측모델 및 이를 적용한 폐암검진 결과를 소개하고자 한다.

위험예측모델이란?

위험예측모델은 특정 사건의 위험을 개별 환자에 적용하여 계산해 낼 수 있는 여러 예측 인자의 조합으로 정의할 수 있다. 위험예측모델은 임상 의사결정(clinical decision making process)에 사용되며, 정보에 근거한 치료 선택에 도움을 준다. 환자 특성 기반의 예측 변수들을 결과 예측에 사용하는 위험예측모델은 임상 의학 발전에 큰 역할을 하고 있다. 개발된 개별 예측모델이 임상적 유용성을 지니기 위해서는 일단 모델이 정확하고 믿을만해야 하며, 대상 집단에 일반화가 가능해야 한다. 그러나, 현실적으로는 개발된 예측모델이 모델 개발에 사용된 대상군 외에서는 제대로 작동하지 않는 경우가 적지 않으며, 이는 연구 설계의 오류나 연구 대상의 특성 차이, 다양한 측정 방법 등에 기인한다(7, 8). 예를 들어, 예측 인자의 수와 비교하여 결과(outcome)의 수가 적은 데이터를 사용하여 개발된 모델은 새로운 환자군에 적용될 때 낮은 예측능을 보일 수 있다. 이는 통계학적으로 모델의 “과적합(overfitting)”으로 설명된다. 과적합 예측모델은 위험도가 낮은 환자의 경우 위험도를 과소평가하고, 위험도가 높은 환자의 경우 과대평가하는 경향이 있어 임상 의사결정에 영향을 줄 수 있다. 따라서, 대상수가 크고, 특성이 잘 기술된, 다양한 집단에서 예측모델의 외부 검증(external validation)을 시행하는 것이 필수적이다(9).

모델의 예측능은 전통적으로 구분(discrimination)과 보정(calibration)의 측정을 통해 평가된다(10). 구분(discrimination)이란 모델이 대상을 양성(positive)과 음성(negative)으로 구별하는 능력을 의미하며, 일반적으로 생존분석 Cox 모델의 일치도 통계(C-statistics 또는 C-index)나 수신자 작동특성곡선(receiver operating characteristics curve; 이하 ROC curve) 분석의 곡선하면적(area under the curve; 이하 AUC)으로 표현된다. C-statistic은 무작위로 선택된 환자가 사건을 경험한 무작위로 선택된 환자보다 모델에서 더 높은 점수를 받는 사건을 경험할 기회를 의미한다. 예를 들어, C-statistics 0.5는 모델이 무작위 선택보다 결과 예측에 우월하지 않음을 의미하며,

1.0은 완벽한 모델을 의미한다(10). 이진법(binary) 결과의 경우 C-statistics는 AUC와 같은 의미를 지니며, ROC curve가 클수록 성능이 좋음을 의미한다. Hosmer 등(11)에 의하면 AUC 0.5~0.7은 poor, 0.7~0.8은 acceptable, 0.8~0.9는 excellent로 간주된다. 보정(calibration)은 모델이 예측한 결과(E)와 실제 관측된 결과(O)가 얼마나 일치하는지를 의미하며, E/O 값이 1.0에 가까울수록 교정이 잘 된 모델임을 의미하며, 1.0 미만일 때는 과소적합(underfitting), 1.0 초과 시에는 과적합(overfitting)을 의미한다. 보정은 흔히 보정 플롯(calibration plot)을 통해 시각화되거나 Hosmer-Lemeshow 테스트로 정량화된 적합도(goodness of fit)로 평가된다($p < 0.05$ 는 보정 불량을 의미). 그밖에 모델의 성능을 측정하는 방법에는 민감도, 특이도, 양성 예측치, 음성 예측치, 정확도, Brier 점수, R2 statistics, 재분류표(reclassification table) 또는 순재편성 개선(net-reclassification improvement), decision curve analysis 등이 있다(12).

모델의 유효성을 평가하기 위해서, 내부(재현성) 및 외부(일반화 가능성) 검증을 위한 다양한 방법을 사용할 수 있다. 가장 보편적으로 사용되는 내부 검증 기법에는 명백한 검증(apparent validation: 동일한 모집단에서의 개발 및 검증), 분할-표본 검증(split-sample validation: 훈련 및 테스트 세트로 분류된 코호트), 교차-검증(cross-validation: 코호트 일부에서 개발 후 나머지 코호트에서 검증 및 상기 과정의 반복), 그리고 부트스트래핑(bootstrapping: 같은 크기의 무작위 표본교체)이 있다. 외부 검증 방법에는 지리적 검증(geographical validation: 다른 센터 또는 국가의 코호트 이용), 일시적 유효성 확인(temporal validation: 동일한 센터 내에서 다른 시간 간격으로 처리한 환자) 또는 완전히 독립적인 유효성 확인(fully independent validation: 다른 기관의 다른 연구자에 의한 검증) 등이 있다(10, 12, 13). 좋은 예측모델은 구분과 보정 모두 우수해야 하며, 구분이 우수한 모델이더라도 보정이 잘못된 경우에는 위험도 예측에 오류가 발생하므로 사용해서는 안 된다(14). 저선량 CT 폐암검진의 고위험군 선정에 우수한 성능을 지닌 위험예측모델을 적용할 수 있다면, 현재의 검진 기준과 비교하여 보다 빨리 초기 폐암을 발견하는 동시에 검진 대상자 수와 위양성을 감소 시켜, 검진의 이득과 위해의 균형을 최적화할 수 있을 것이다.

전통적 위험인자를 포함하는 예측모델

Bach 모델

Bach 모델은 가장 처음 개발된 폐암 위험예측모델이다. 이 모델은 카로틴과 레티놀의 효능을 연구하는 Carotene and Retinol Efficacy Trial 연구에서 흡연 고위험군을 대상으로 폐암 발생의 1년 위험도와 폐암 사망 위험률을 예측하는 모델로 개발되었다(15). 이 모델은 18172명의 대상군에서 연령(50~75세), 성별, 흡연기간, 흡연강도(하루 10~60개비), 금연기간(20년 이내), 석면 노출력을 모델의 변수로 설정하였다. 대상군에는 14000여 명의 과흡연군(heavy smoker)과 4000여 명의 석면 노출군이 포함되었다. 내부 검증(internal validation)에서 AUC 0.72와 양호한 구분(discrimination)을 보였다. 이 모델에 대하여 6239명을 대상으로 한 외부 검증이 핀란드에서 시행되었다(16). 이 연구에서는 폐암 예측치와 실제 발생 값을 비교하여 모델 성능을 분석하였는데, 10년간의 폐암 발생 중 89%를 예측하여 모델이 실제 위험을 11%가량 과소평가하는 경향이 있음이 제한점으로 지적된다(16).

Liverpool Lung Project 모델

Liverpool Lung Project (이하 LLP) 모델은 2008년 영국 리버풀 지역에서 40~80세의 1736예의 환자-대조군을 대상으로 개발되었다(17). 이 모델에서는 흡연기간, 석면 노출력, 폐렴, 암에 대한 개인 및 가족력을 바탕으로 5년 폐암 위험률을 계산하였다. 이 모델은 양호한 구분(AUC 0.70)을 보였는데, 2.5% 위험률을 문턱값으로 정했을 때 모델의 성능이 가장 우수하였으며 이는 62%의 민감도와 70%의 특이도에 해당되었다. Cassidy 등(17), Raji 등(18), D'Amelio 등(19)에 의해 LLP 모델에 대한 내부 및 외부 검증이 시행되었는데, 2.5% 위험률 문턱값에서 Youden index 0.32~0.417로, LLP 모델의 최적 문턱값은 2.5% 임을 의미하였다. 5% 문턱값에서 LLP 모델은 80% 이상의 대조군을 검진 대상에서 배제할 수 있었으나, 민감도가 낮아서 35.7~57.4%의 폐암만을 찾을 수 있었다. 10% 위험률 문턱값에서는 90% 이상의 대조군을 검진 대상에서 배제할 수 있었으나, 14~27%의 폐암만 찾을 수 있었다. LLP 모델은 이후 United Kingdom Lung Screening Trial에 적용되어, 50~75세의 폐암 고위험군 대상의 저선량 CT 폐암검진에 적용되고 있다(5년 동안 5% 이상의 문턱값) (20). 이 연구에서 고위험률 문턱값과 1회 CT 검진을 채택한 것은 검진의 효율과 비용효과를 최대화하기 위함이다.

Spitz 모델

Spitz 모델은 2007년 2889명의 폐암 환자-대조군 연구에서 파생되어, 역학 및 임상 데이터(Surveillance, Epidemiology, and End Results; 이하 SEER)의 추가를 통해 고안되었다(21). 예측모델의 성능에 영향을 줄 수 있는 젊은 폐암 환자(50세 미만)들은 대상군에서 제외되었다. Spitz 모델은 1년간의 폐암 절대 위험률을 예측하기 위해 고안되었으나, 원하는 년 수까지 기간을 연장하기 위해 반복적으로 실행될 수 있다. Spitz 모델은 비흡연자, 과거 흡연자 및 현재 흡연자에 대한 개별 로지스틱 회귀모델을 적용하며, 흡연 여부, 가족 암 병력, 석면 및 먼지 노출, 이전의 호흡기 질환, 꽃가루 알레르기 병력 및 흡연력 변수가 포함되었다. Spitz 모델의 장점은 포함된 변수들이 생물학적 타당성이 높고, 환자 인터뷰를 통해 확인하기 쉽다는 점이다. 그러나, Spitz 모델의 과거 및 현재 흡연자 모델에 대한 실제 AUC 통계는 다른 위험예측모델들과 유사한 결과(modest, AUC 0.63 및 0.58)를 보였다(21).

African American 모델

미국이나 서유럽에서 개발된 위험예측모델의 대부분은 인종(ethnicity)이 폐암 발생에 미치는 영향을 고려하고 있지 않다. 2008년 Etzel 등(22)은 아프리카계 미국인의 폐암 예측 로지스틱 회귀모델에서 Spitz 모델과 유사한 5년 절대 폐암 위험률을 계산하였다. 이 모델은 인종을 위험예측모델에 포함하였다는 데 의미를 부여할 수 있다. 인종을 제외한 다른 선정 고려사항은 없었으며, 나이, 성별, 인종이 보정되었다. SEER 자료에서 20세 이상 아프리카계 미국인의 폐암 발생률과 사망률을 이용하였다. 988명의 내부 검증을 통해 모델이 개발되었으며(AUC 0.75), 156명에서의 내부 검증에서 AUC 0.75, 325명의 외부 검증에서 AUC 0.63을 보였다(22).

Prostate, Lung, Colorectal, and Ovarian 모델

2011년 Tammemägi 등(23)은 전립선암, 폐암, 대장암, 난소암 선별 검사(Prostate, Lung, Colorectal, and Ovarian; 이하 PLCO)에서 흡연자와 비흡연자를 모두 포함하는 첫 번째 위험예측 모델을 구축하였다. 흡연력에 무관한 모델 1과 흡연자 대상의 모델 2는 9년 간의 폐암 위험을 예측하며, 두 모델 모두 연령, 사회경제적지위(교육), 체질량지수, 폐암 가족력, 만성폐색성폐질환(chronic obstructive pulmonary disease; 이하 COPD), 최근 흉부 X선, 흡연 상태, 흡연 갑년 및 흡연기간이 포함되었고, 모델 2에는 금연기간이 포함되었다. 대조군 피험자의 추적 관찰기간(중앙값 9.2년) 동안 1040건의 폐암이 발생하였고, 외부 검증의 추적 관찰 중에 213건의 폐암이 발생하였다. 모델 1과 모델 2의 AUC는 각각 0.857과 0.805이고, 외부 검증에서 모델 1과 모델 2는 각각 0.841과 0.784의 AUC를 보였다(23). 또한, 여성, 남성, 백인 및 비백인에서 높은 구분(discrimination)을 보였다. 이 모델은 우수한 예측력을 보여 폐암 위험예측모델로서의 성능을 인정받았으며, 추후 PLCO_{M2012} 모델로 수정되었다.

PLCO_{M2012} 모델

앞서 기술한 것처럼 PLCO 모델은 비교적 높은 예측력을 보였으나, 모델링 과정이 복잡하고 9.2년간의 추적기간을 통해 위험도가 계산되므로 NLST에 직접 적용하기는 어렵다. 이에 PLCO 모델의 NLST에서의 예측력을 평가하기 위하여, PLCO 연구에 포함되었던 80375명의 환자-대조군에서 6년 폐암 위험률 계산을 위한 PLCO_{M2012} 모델이 개발되었다(24). 2011년 PLCO 모델의 예측 인자였던 흡연 갑년(pack years)은 각 성분의 효과를 더 잘 이해할 수 있도록 흡연 강도와 흡연기간으로 세분화되었다. PLCO_{M2012}의 예측 인자에는 연령, 인종/민족, 사회경제적지위, 신체질량지수, 암 병력, 폐암 가족력, COPD 병력, 흡연 상태(이전, 현재), 흡연강도, 흡연기간 및 금연 후 기간이 포함된다. 이 모델은 고위험군 NLST 코호트와 55~74세의 30갑년 이상의 흡연력 및 금연 15년 이내의 PLCO 코호트에서 검증되었다. PLCO_{M2012} 모델은 폐암 위험률과 관련이 없는 것으로 밝혀진 흡연 갑년과 흉부 X선은 고려하지 않았다. 모델의 AUC는 개발 데이터세트에서 0.803, 검증 데이터세트에서 0.797이었다. NLST 기준과 비교하여, PLCO_{M2012} 기준은 우수한 민감도(83.0% vs. 71.1%, $p < 0.001$)와 양성예측도(4.0% vs. 3.4%, $p = 0.01$)를 보이면서도 특이도는 감소되지 않았다(62.9% vs. 62.7%, $p = 0.54$). PLCO_{M2012} 모델은 NLST와 비교하여 678개의 폐암 중에서 81개의 폐암을 더 발견하여 NLST 대비 폐암을 놓칠 확률은 41.3%에 불과하였다. 이러한 결과는 NLST 기준을 그대로 적용하는 것보다 정확한 위험예측모델을 사용하는 것이 폐암검진 대상 선정에 더 효율적이라는 것을 의미한다.

Tammemägi 등(25)은 PLCO_{M2012} 모델과 NLST 사망률 결과를 토대로 위험 임계치(risk threshold)를 제시하였다. PLCO_{M2012} 위험도 ≥ 0.0151 (65 percentile)에서는 대조군(chest X-ray)과 비교하여 NLST low-dose CT 검진군에서 폐암 사망률이 일관되게 낮았다. 1명의 폐암 사망을 예방하기 위해 필요한 검진의 수는 65~100 percentile 위험군에서 255 [95% confidence interval (이하 CI), 143~1184], 30~65 percentile 위험군은 963 (95% CI, 291~754)이었고, 30 percentile 위험군에서는 폐암 사망이 없어서 값이 계산되지 못했다. USPSTF 기준과 비교하여 PLCO_{M2012} 위험도

≥ 0.0151 기준을 PLCO 중재군에 적용했을 때 양성결과는 8.8% 더 적었고($p < 0.001$), 12.4% 더 많은 폐암을 발견했다(민감도 80.1% vs. 71.2%, $p < 0.001$). 또한, 위양성이 적었고(특이도 66.2% vs. 62.7%, $p < 0.001$), 양성예측도가 더 높았다(4.2% vs. 3.4%, $p < 0.001$). USPSTF 기준을 적용할 때 PLCO_{m2012} 기준으로는 저위험군으로 간주되는 경우가 상당수 포함되었는데, USPSTF 대상자의 26%는 PLCO_{m2012} 위험도 ≥ 0.0151 이하에 해당되었다. NLST와 USPSTF 기준 적용 시 고위험군의 일부는 대상에서 제외되었는데, 금연 15년 이상의 PLCO 환자 8.5%에서 PLCO_{m2012} 위험도가 ≥ 0.0151 이었고, 이 중 2.9%는 6년 추적검사에서 폐암으로 진단되었다. 이 연구를 통해 Tammemägi 등(25)은 폐암검진 대상자의 선정에 NLST 기준보다 PLCO_{m2012} 모델 및 PLCO_{m2012} 위험도 ≥ 0.0151 임계 값을 사용하는 것이 바람직하다고 주장하고 있다.

Hoggart 모델

Hoggart 모델에서는 과거 및 현재 흡연자를 위한 Weibull parametric 생존 모델을 위해 European Prospective Investigation into Cancer and Nutrition 연구의 전향적 코호트 데이터를 사용하였다(26). 169035명의 흡연자가 포함되었고, 비흡연자는 Hoggart 모델에서 제외되었다. 이 모델에서는 단일 뉴클레오타이드 다형성(single nucleotide polymorphisms; 이하 SNPs) (rs8034191/15q25 및 s402710/5p15)이 평가되었으나, 폐암과의 연관성은 발견되지 않았다. 석면, 실리카, 금속 및 다환 방향족 탄화수소 관련 직업적 노출력은 평가되었으나 유용하지 않았다. 최종 모델 예측 인자에는 연령, 흡연 강도, 흡연 지속 기간, 흡연 중단 후 년수 및 흡연 개시 연령이 포함되었다. 현재 흡연자(current smoker), 과거 흡연자(former smoker) 및 전체 흡연자(ever-smoker)의 경우, 검증 데이터에서의 5년 위험 AUC는 각각 0.767 (95% CI, 0.701~0.832), 0.715 (95% CI, 0.532~0.898) 및 0.787 (95% CI, 0.728~0.847)이었다. 전반적으로 Hoggart 모델은 구분(discrimination)과 보정(calibration)이 우수하였다. Hoggart 모델은 폐암검진을 위한 대상 선정에 잠재적 유용성을 보였으며, 추가 데이터를 이용한 외부 검증이 필요하다.

임상적, 분자생물학적, 유전학적 위험인자를 포함하는 예측모델

임상적, 분자생물학적, 유전학적 위험인자를 포함하는 예측모델이 적용되기 위해서는 의사의 진찰이 필요하며, 검체를 얻거나 추가 검사를 시행해야 하므로, 이에 따른 의료비용의 상승과 추가 시간 및 노력이 필요하다. 이러한 부가적인 노력에도 불구하고 개발된 예측모델이 실제로 적용되려면 기존 모델보다 예측력이 우수해야 한다.

Extended Spitz 모델

Spitz 등(27)은 기존 모델(18)에 2개의 생물학적 인자(DNA 복구능력과 돌연변이 유발원 bleomycin에 대한 민감도)를 포함하는 확장 모델을 개발하였다. 과거 흡연자와 현재 흡연자에서 교차 확인된 AUC는 0.68과 0.70이었다. 이 모델에 포함된 생체표지자는 모델의 AUC 향상을 가져왔지만 개선의 크기는 크지 않았으며, 이 모델의 구분(discrimination)도 다른 위험예측모델에 비해

상대적으로 낮아 임상적 유용성은 제한적으로 평가된다.

유전자 기반 위험 점수(Gene-Based Risk Score)

Young 등(28)은 감수성(susceptibility)이나 보호(protection)의 유전적 표지자로서 20개의 SNP를 포함하는 예측모델을 만들기 위해 2단계의 환자-대조군 연구를 설계하였다. 446명의 폐암 환자와 484명의 건강한 흡연자를 대상으로 모델의 추가 예측 인자에는 연령, COPD 및 폐암의 가족력이 포함되었고, COPD는 폐기능 검사로 평가되었다. 환자와 대조군은 흡연력으로 매칭되어, 폐암의 가장 중요한 위험인자인 흡연은 예측 알고리즘에서 제외되었다. 이 모델은 일반인을 대상으로 사용하기 위해 보정은 하지 않았다. 제시된 대표적인 AUC는 0.75로 비교적 우수하였으나, 대조군에는 COPD가 배제되고 환자군에는 COPD가 허용된 비표준 설계(nonstandardized design) 형태를 취하고 있다. 기타 연구 제외 기준은 환자군 및 대조군에 유사하게 적용되었다. 잘 알려진 것처럼 COPD는 흡연과 관련이 있고, 따라서 폐암의 독립적인 위험 인자이다. 이러한 비표준 설계를 택한 모델은 AUC를 상향 편향시킬 가능성이 있으므로 이러한 단점을 결과 해석에 고려해야 한다.

한국 남성(Korean Men) 모델

Park 등(29)은 한국 남성에 대한 전향적 자료를 사용하여 8년간의 폐암을 예측하는 Cox 회귀모델을 개발하였다. 모델 예측 인자에는 흡연 노출, 흡연 개시 연령, 체질량지수, 신체 활동 및 공복 혈당 수준이 포함되었다. 외부 검증 데이터에서 모델의 C-statistics는 0.871 (95% CI, 0.867~0.876) 이었고, 분석에 비흡연자를 포함하면 C-statistics가 크게 증가된다. 이 연구의 일부 예측 변수의 모델링은 다음과 같은 제한점이 있다. 예를 들어 금연기간이 15년 이전과 이후로 이분화되었고, 흡연 강도는 현재 흡연자의 범주로만 모델링 되었다. 또한, 모델링 특성상, 이 모델이 여성이나 비아시아인에게 일반화되어 적용되기는 어려운 단점이 있다. 그리고, 이 모델은 공복 혈당치 정보를 필요로 하는데, 이는 일반 대중에서 쉽게 얻을 수 있는 자료는 아니다. 이러한 모델링 차원의 제약에도 불구하고 이 모델은 미국이나 유럽 외에서 개발된 단 2개의 위험예측모델 중 하나로서, 130만 명 이상의 비교적 큰 집단에서 개발 적용되었다는데 의미를 부여할 수 있다.

Pulmonary Function with Lung Response 모델

Tammemagi 등(30)은 폐기능(forced expiratory volume)과 객담 DNA 영상 계측법이 포함된 예측모델을 개발하였다. 이 모델은 20갑년 이상의 40세 이상 폐암 고위험군 2596명을 대상으로 하였다. 나이, 성별, 체질량지수, 교육 수준, 폐암 가족력, 흡연력이 모델 설계에 포함되었으며, 2522명의 환자를 대상으로 내부적으로 검증되었다. 폐기능 감소와 DNA 영상에서의 세포 수 증가가 폐암 위험 증가와 연관이 있었으며, 흡연자에게 적용할 때 좋은 교정($p=0.313$) 및 AUC (0.773)를 보였다.

Chinese Multifactorial Genetic 모델

Li 등(31)은 중국에 거주하는 5068명의 환자-대조군 자료를 이용하여 로지스틱 회귀 위험 모델을

개발하였다. 이 모델에는 5개의 SNP 유전자 데이터가 점수화되어 포함되었다. SNP 점수와 흡연 상태를 포함한 모델은 AUC 0.639 (95% CI, 0.621~0.652)였고, 주로 흡연 상태에 의해 결정되었다. 이 모델은 흡연을 적절히 모델링하지 못하였고 다른 유용한 예측 인자들이 제외되어 검진을 위한 적절한 대상을 선별하기에는 불충분한 것으로 평가된다.

Hippisley-Cox와 Coupland 모델

Hippisley-Cox와 Coupland (32)는 전향적(prospective) 영국 1차 의료(primary care) 자료를 기반으로 2년 내 폐암 발생에 대한 Cox 회귀 예측모델을 만들었다. 모델 개발에 2406127명, 검증에 1267151명이 포함되었고, 남성과 여성을 위한 별도의 모델이 개발되었다. 나이는 모델에 기본 예측 시간함수로 포함되었지만 별도의 예측 변수로는 포함되지 않았다. 흡연 상태는 흡연 강도(현재 흡연자에게만 사용 가능)와 함께 합쳐졌다. 추가 예측 인자로 사회경제적지위, 과거 암 병력, 체질량지수, COPD, 객혈, 식욕 감퇴, 체중 감소, 새로 생긴 기침 및 헤모글로빈 농도(< 11 g/dL)가 포함되었다. 검증 데이터에서 남, 여 모델의 AUC는 각각 0.92 (95% CI, 0.91~0.93)로 상당히 높았는데, 분석에 비흡연자를 포함한 것이 AUC 값이 상승하는 주요 원인으로 추정된다. 이 모델은 추적 기간이 짧아 장기간의 추적 관찰기간 동안 높은 예측 성능을 유지하기 어려운 것이 단점으로 지적된다.

Iyen-Omofoman 모델

Iyen-Omofoman 등(33)은 영국의 일반의(general practitioner) 진료 자료를 사용하여 연령, 성별, 사회경제적지위 및 흡연 강도와 흡연 상태를 포함하는 로지스틱 회귀 예측모델을 만들었다. 2년 내 폐암 진단을 받은 군($n = 12074$)과 대조군($n = 120731$)을 비교하였다. 진단 4~12개월 전에 수집된 임상 예측인자에는 기침, 객혈, 호흡 곤란, 체중 감소, 하부 호흡기 감염, 비특이적 흉부 감염 및 COPD가 포함되었다. 검증 데이터의 AUC는 0.88이었으나, 분석에 비흡연자를 포함시킨 것이 AUC 값이 상승하는 원인으로 추정된다. 이 연구에서의 보정 통계는 제공되지 않은 것이 단점으로 지적되며, 일반인의 폐암 위험 평가를 통해 만들어진 모델이므로, 폐암 고위험군의 검진 대상 선정에는 유용하지 않은 것으로 평가된다.

위험예측모델의 저선량 CT 폐암검진 적용

미국의 USPSTF나 Centers for Medicare and Medicaid Services (이하 CMS)는 아직 폐암 위험모델에 기반한 검진 대상자 선택을 권고하고 있지는 않다. NCCN 2018 권고안이 처음으로 PCLO_{m2012} 위험모델에 기반한 폐암검진 대상자 선정을 권고하였다. 현재까지, NLST 기준과 비교하여 모델-기반의 위험도 예측이 폐암 발견의 민감도와 양성예측도 및 비용효과 측면에서 더 우월하다는 증거들이 다수 제시되고 있다. 그러나, 이러한 근거들은 주로 기존 검진 자료의 후향적 분석이나 시뮬레이션을 통해 얻어졌다는 점이 제한점으로 지적된다. 최근, 전향적 연구를 통한 위험예측 모델의 폐암검진 적용이 진행되고 있다. International Lung Screen Trial (이하 ILST)에서는 캐나다, 호주 등에서 4천여 명의 참가자를 전향적으로 모집하고 있으며, PCLO_{m2012} 6년 위험도 1.5% 이

상 또는 USPSTF 기준을 적용하고 있다(34). ILST의 연구 설계는 흥미로운데, PCLO_{m2012} 및 USPSTF 기준이 모두 양성 또는 음성일 때에는 당연히 검진 대상에 포함되거나 제외되며, PCLO_{m2012} 양성/USPSTF 음성이거나 PCLO_{m2012} 음성/USPSTF 양성처럼 두 기준이 서로 상이한 경우 추적 관찰을 통하여 자연스럽게 두 모델의 성능을 비교해 볼 수 있다. 또한, USPSTF 기준은 NLST와 CMS 기준을 기반으로 하고 있으므로 이들 기준에 대한 비교도 가능할 것으로 보인다. 캐나다의 Cancer Care Ontario에서는 3000명을 대상으로 폐암검진을 위한 시범사업을 진행하고 있다(35). 검진 대상자는 다학제 전문가 패널에 의해 PCLO_{m2012} 위험도 2% 이상으로 선정하였으며, 2020년 최종 결과를 보고할 예정이다. 이 연구를 통해 위험예측모델의 현실적 적용과 위험예측모델이 조기 폐암으로 진단될 대상자를 성공적으로 선택할 수 있는지에 대한 유용한 정보를 제공할 수 있을 것으로 기대된다. 앞으로 보다 많은 수의 전향적 연구를 통해서 위험예측모델의 우수성이 추가로 증명된다면, 향후 검진 대상 선정에 있어 모델 기반의 위험도를 적용하는 것이 보편화될 수 있을 것으로 예상된다.

폐암검진의 고위험 대상군을 선정하기 위한 예측모델 외에도, 저선량 폐암검진 CT에서 발견된 폐결절의 악성 위험도 예측을 위한 모델 개발도 활발히 진행 중이다. 캐나다의 McWilliams 등(36)은 저선량 CT 검진 자료를 이용하여 발견된 폐결절의 악성종양 가능성을 예측하는 위험도 계산기(Vancouver Risk Calculator, Brock 모델 또는 McWilliams 모델로도 불림)를 개발하였다. 개발 데이터세트에는 Pan-Canadian Early Detection of Lung Cancer Study 연구 자료가 사용되었고, 검증에는 British Columbia Cancer Agency 화학예방 검진(chemoprevention trial) 자료를 사용하여, 기저 저선량 CT에서 발견된 결절의 악성도 계산을 위한 다변량 로지스틱 회귀모델이 개발되었다. 이 모델에서 악성의 예측인자는 고령, 여성, 폐암의 가족력, 폐기종, 큰 결절, 우상엽의 결절, 부분고형결절(part-solid nodule), 적은 수의 결절, 침상형 결절(spiculated nodule)이었다. 이 모델은 10 mm 이하의 결절을 포함하는 검증 데이터세트에서 AUC 0.9 이상의 우수한 구분(discrimination) 및 보정(calibration)을 보였다. Brock 모델을 NLST 코호트에 적용한 White 등(37)의 연구에서, 이 모델은 악성 예측에 있어 높은 성능을 보였으나(민감도 85.3%, 특이도 93.9%, 음성예측도 99.6%), 위험점수 문턱값을 10%로 정의했을 때의 양성예측도는 27.4%로 다소 낮았으며, 상위 9개 위험인자가 모두 적용 가능한 NLST 대상군에서는 AUC 0.963으로 매우 우수한 성능을 보였다. 동일 연구자들은 최근 Brock 모델과 Lung Imaging Reporting and Data System (이하 Lung-RADS)의 악성 예측을 NLST 자료를 대상으로 비교하였다(38). 총 4408개의 결절 중에서 4078개의 고형 결절 중 100개(2.5%)가 폐암이었고, 330개의 부분고형 결절 중 10개(3%)가 폐암이었다. 악성 예측의 민감도, 특이도, 정확도는 결절 기반으로 비교 시 Brock 모델(93%, 90%, 90%)이 Lung-RADS (87%, 83%, 83%)에 비해 우수하였고, 환자 기반으로 비교 시에도 Brock 모델(93%, 85%, 85%)이 Lung-RADS (87%, 76%, 76%)보다 우수하였다.

부분고형결절을 침습폐선암(invasive pulmonary adenocarcinoma)과 침습전폐선암(preinvasive adenocarcinoma)으로 구별하는 것은 잠재적 폐암으로서 폐결절의 적절한 관리에 있어 중요한 의미를 지닌다. 결절 크기, 고형 부분(solid proportion), 결절 유형 등 폐결절의 세 가지 특성만을 이용한 Lee 모델은 9개의 특성을 이용한 Brock 모델보다 침습폐선암의 진단에 좋은 성능을 보

였다(39, 40). 향후 Lung-RADS의 개정에 폐암 예측모델에 포함된 것처럼 악성을 시사하는 결절 및 환자 특성이 반영된다면, 폐암 진단에 큰 도움이 될 것으로 기대된다.

결론

NLST가 저선량 CT를 이용한 폐암검진을 통해 폐암 고위험군에서 20%의 사망률 감소를 가져올 수 있다고 보고한 이래, 폐암검진 분야는 매우 빠른 속도로 발전하고 있다. 그중에서도 폐암 고위험군을 정밀하게 정의하고 적절한 검진 대상자를 선택하는 것은 폐암검진의 비용대비 효과를 극대화하기 위하여 매우 중요한 분야이며 많은 연구가 진행되고 있다. 우리나라에서도 2019년 하반기부터 폐암검진이 국가암검진의 일환으로 시행될 예정이다. 앞서 기술된 것처럼 폐암의 예측모델에는 흡연, 성별, 연령, 인종, 직업력, 사회경제적지위 등 다양한 요인들이 포함될 수 있다. 우리나라를 포함하는 아시아 국가에는 서양과 비교하여 특히 간유리음영 또는 부분고형결절 형태의 폐선암의 발생 빈도가 높는데, 폐암 발생 과정의 인종, 유전적 소인 및 흡연에 대한 감수성 차이가 그 원인으로 추정된다(41, 42). 따라서, 서양에서 개발된 폐암예측모델이 인종과 폐암 감수성에서 차이가 예상되는 우리나라에서도 우수한 성능을 보일 수 있을지는 미지수이다. 향후 우리나라 상황에 맞는 우수한 폐암 위험예측모델의 개발을 통해 폐암 고위험군을 적절히 선별하고 폐암 사망률을 감소시킬 수 있기를 기대해 본다.

Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

REFERENCES

1. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395-409
2. Moyer VA; U.S. Preventive Services Task Force. Screening for lung cancer: US Preventive Services Task Force recommendation statement. *Ann Intern Med* 2014;160:330-338
3. Jacobson FL, Austin JH, Field JK, Jett JR, Keshavjee S, MacMahon H, et al. Development of The American Association for Thoracic Surgery guidelines for low-dose computed tomography scans to screen for lung cancer in North America: recommendations of The American Association for Thoracic Surgery Task Force for Lung Cancer Screening and Surveillance. *J Thorac Cardiovasc Surg* 2012;144:25-32
4. Wender R, Fontham ET, Barrera E Jr, Colditz GA, Church TR, Ettinger DS, et al. American Cancer Society lung cancer screening guidelines. *CA Cancer J Clin* 2013;63:106-117
5. Wood DE, Eapen GA, Ettinger DS, Hou L, Jackman D, Kazerooni E, et al. Lung cancer screening. *J Natl Compr Canc Netw* 2012;10:240-265
6. Bach PB, Mirkin JN, Oliver TK, Azzoli CG, Berry DA, Brawley OW, et al. Benefits and harms of CT screening for lung cancer: a systematic review. *JAMA* 2012;307:2418-2429
7. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med* 1999;130:515-524
8. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;338:b605
9. Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol* 2003;56:826-832
10. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med* 2000;19:453-473

11. Hosmer Jr DW, Lemeshow S, Sturdivant RX. *Applied logistic regression*. Hoboken: John Wiley & Sons 2013
12. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012;98:691-698
13. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *BMJ* 2009;338:b604
14. Sakoda LC, Henderson LM, Caverly TJ, Wernli KJ, Katki HA. Applying risk prediction models to optimize lung cancer screening: current knowledge, challenges, and future directions. *Curr Epidemiol Rep* 2017;4:307-320
15. Bach PB, Kattan MW, Thornquist MD, Kris MG, Tate RC, Barnett MJ, et al. Variations in lung cancer risk among smokers. *J Natl Cancer Inst* 2003;95:470-478
16. Cronin KA, Gail MH, Zou Z, Bach PB, Virtamo J, Albanes D. Validation of a model of lung cancer risk prediction among smokers. *J Natl Cancer Inst* 2006;98:637-640
17. Cassidy A, Myles JP, Van Tongeren M, Page RD, Liloglou T, Duffy SW, et al. The LLP risk model: an individual risk prediction model for lung cancer. *Br J Cancer* 2008;98:270-276
18. Raji OY, Duffy SW, Agbaje OF, Baker SG, Christiani DC, Cassidy A, et al. Predictive accuracy of the Liverpool Lung Project risk model for stratifying patients for computed tomography screening for lung cancer: a case-control and cohort validation study. *Ann Intern Med* 2012;157:242-250
19. D'Amelio AM Jr, Cassidy A, Asomaning K, Raji OY, Duffy SW, Field JK, et al. Comparison of discriminatory power and accuracy of three lung cancer risk models. *Br J Cancer* 2010;103:423-429
20. McDonald FE, Yadekarfar G, Baldwin DR, Devaraj A, Brain KE, Eisen T, et al. The UK Lung Screen (UKLS): demographic profile of first 88,897 approaches provides recommendations for population screening. *Cancer Prev Res (Phila)* 2014;7:362-371
21. Spitz MR, Hong WK, Amos CI, Wu X, Schabath MB, Dong Q, et al. A risk model for prediction of lung cancer. *J Natl Cancer Inst* 2007;99:715-726
22. Etzel CJ, Kachroo S, Liu M, D'Amelio A, Dong Q, Cote ML, et al. Development and validation of a lung cancer risk prediction model for African-Americans. *Cancer Prev Res (Phila)* 2008;1:255-265
23. Tammemägi MC, Pinsky PF, Caporaso NE, Kvale PA, Hocking WG, Church TR, et al. Lung cancer risk prediction: prostate, lung, colorectal and ovarian cancer screening trial models and validation. *J Natl Cancer Inst* 2011;103:1058-1068
24. Tammemägi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, et al. Selection criteria for lung-cancer screening. *N Engl J Med* 2013;368:728-736
25. Tammemägi MC, Church TR, Hocking WG, Silvestri GA, Kvale PA, Riley TL, et al. Evaluation of the lung cancer risks at which to screen ever- and never-smokers: screening rules applied to the PLCO and NLST cohorts. *PLoS Med* 2014;11:e1001764
26. Hoggart C, Brennan P, Tjonneland A, Vogel U, Overvad K, Østergaard JN, et al. A risk model for lung cancer incidence. *Cancer Prev Res (Phila)* 2012;5:834-846
27. Spitz MR, Etzel CJ, Dong Q, Amos CI, Wei Q, Wu X, et al. An expanded risk prediction model for lung cancer. *Cancer Prev Res (Phila)* 2008;1:250-254
28. Young RP, Hopkins RJ, Hay BA, Epton MJ, Mills GD, Black PN, et al. A gene-based risk score for lung cancer susceptibility in smokers and ex-smokers. *Postgrad Med J* 2009;85:515-524
29. Park S, Nam BH, Yang HR, Lee JA, Lim H, Han JT, et al. Individualized risk prediction model for lung cancer in Korean men. *PLoS One* 2013;8:e54823
30. Tammemägi MC, Lam SC, McWilliams AM, Sin DD. Incremental value of pulmonary function and sputum DNA image cytometry in lung cancer risk prediction. *Cancer Prev Res (Phila)* 2011;4:552-561
31. Li H, Yang L, Zhao X, Wang J, Qian J, Chen H, et al. Prediction of lung cancer risk in a Chinese population using a multifactorial genetic model. *BMC Med Genet* 2012;13:118
32. Hippisley-Cox J, Coupland C. Identifying patients with suspected lung cancer in primary care: derivation and validation of an algorithm. *Br J Gen Pract* 2011;61:e715-e723
33. Iyen-Omofoman B, Tata LJ, Baldwin DR, Smith CJ, Hubbard RB. Using socio-demographic and early clinical features in general practice to identify people with lung cancer earlier. *Thorax* 2013;68:451-459
34. U.S. National Library of Medicine. International lung screen trial (ILST). Available at: <https://clinicaltrials.gov/ct2/show/NCT02871856>. Accessed Apr 6, 2019
35. Cancer Center Ontario. Available at: <https://www.cancercareontario.ca/en/types-of-cancer/lung>. Accessed on Apr 6, 2019

36. McWilliams A, Tammemagi MC, Mayo JR, Roberts H, Liu G, Soghrati K, et al. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med* 2013;369:910-919
37. White CS, Dharaiya E, Campbell E, Boroczky L. The Vancouver lung cancer risk prediction model: assessment by using a subset of the national lung screening trial cohort. *Radiology* 2017;283:264-272
38. White CS, Dharaiya E, Dalal S, Chen R, Haramati LB. Vancouver risk calculator compared with ACR lung-RADS in predicting malignancy: analysis of the national lung screening trial. *Radiology* 2019;291:205-211
39. Kim H, Park CM, Jeon S, Lee JH, Ahn SY, Yoo RE, et al. Validation of prediction models for risk stratification of incidentally detected pulmonary subsolid nodules: a retrospective cohort study in a Korean tertiary medical centre. *BMJ Open* 2018;8:e019996
40. Kim H, Goo JM, Park CM. A simple prediction model using size measures for discrimination of invasive adenocarcinomas among incidental pulmonary subsolid nodules considered for resection. *Eur Radiol* 2019;29:1674-1683
41. Jung KJ, Jeon C, Jee SH. The effect of smoking on lung cancer: ethnic differences and the smoking paradox. *Epidemiol Health* 2016;38:e2016060
42. Ha SY, Choi SJ, Cho JH, Choi HJ, Lee J, Jung K, et al. Lung cancer in never-smoker Asian females is driven by oncogenic mutations, most often involving EGFR. *Oncotarget* 2015;6:5465-5474

폐암검진에서의 위험예측모델

김태정^{1*} · 김혜영² · 구진모^{3,4} · 선주성⁵

폐암 고위험군에서 저선량 전산화단층촬영을 이용한 폐암검진은 폐암으로 인한 사망률을 20%까지 줄일 수 있다. 최근까지 보고된 여러 연구 결과들은 폐암 위험예측모델의 개발이 폐암검진의 주요 발전 중 하나임을 시사한다. 본 기고에서는 위험예측모델을 평가하기 위한 기술적 요구 사항을 검토하고, 지금까지 보고된 주요 폐암 위험예측모델과 이 모델을 적용한 폐암검진 결과를 소개하고자 한다.

¹성균관대학교 의과대학 삼성서울병원 영상의학과,

²국립암센터 영상의학과,

³서울대학교 의과대학 서울대학교병원 영상의학과,

⁴서울대학교 의과대학 의학연구원, 방사선의학연구소,

⁵아주대학교 의과대학 영상의학교실