MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI230999

Real-World Effectiveness of Lung Cancer Screening Using Deep Learning-Based Counterfactual Prediction

Zheng FENG[b], Zhaoyi CHEN[b], Yi GUO[b], Mattia PROSPERI[c], Hiren MEHTA[d], Dejana BRAITHWAITE[c,d], Yonghui WU[a,b], Jiang BIAN[a,b,1]

[a] *Department of Health Outcomes & Biomedical Informatics, College of Medicine, University of Florida, Gainesville, FL, USA*

[b] *Cancer Informatics Share Resource, University of Florida Health Cancer Center, Gainesville, FL, USA*

[c] *Department of Epidemiology, University of Florida, Gainesville, FL, USA*

[d] *Department of Medicine, University of Florida, Gainesville, FL, USA*

[e] *Department of Surgery, College of Medicine, University of Florida, Gainesville, FL, USA*

Abstract. The benefits and harms of lung cancer screening (LCS) for patients in the real-world clinical setting have been argued. Recently, discriminative prediction modeling of lung cancer with stratified risk factors has been developed to investigate the real-world effectiveness of LCS from observational data. However, most of these studies were conducted at the population level that only measured the difference in the average outcome between groups. In this study, we built counterfactual prediction models for lung cancer risk and mortality and examined for individual patients whether LCS as a hypothetical intervention reduces lung cancer risk and subsequent mortality. We investigated traditional and deep learning (DL)-based causal methods that provide individualized treatment effect (ITE) at the patient level and evaluated them with a cohort from the OneFlorida+ Clinical Research Consortium. We further discussed and demonstrated that the ITE estimation model can be used to personalize clinical decision support for a broader population.

Keywords. Causal effect estimation, deep learning, real-world data, counterfactual

1. Introduction

Lung cancer is the leading cause of cancer-related death in the United States, with an estimated 236,740 new cases in 2022, out of which 130,180 patients will die [1]. The five-year survival rate is 56% for localized cases, but early diagnose occurs only in 16% cases [2]. Early identification of lung cancer is critical due to its high public health burden. Clinical guidelines recommended low-dose CT (LDCT)-based lung cancer screening (LCS) for high risk individuals [3]. For example, the USPSTF's 2014 guideline recommends LDCT for those 55-80 years old, with 30-pack-year smoking history [3], and the 2021 update broadens it to those 50-80 years old with 20 pack-year smoking history [4]. However, LDCT use remains low among high-risk heavy smokers who

[1] Corresponding Author: Jiang Bian, PhD; bianjiang@ufl.edu; Address: 2197 Mowry Road, 122 PO Box 100177 Gainesville, FL 32610-0177; Phone Number: +1 (352) 273-8878.

would benefit the most from LCS, and there's overuse among low-risk individuals, where the high false-positive rate associated with LDCT (i.e., 23.3% from the original NLST trial) would pose harm that can lead to postprocedural complications [5]. Prediction models like PLCO2012 help guide LCS decisions. For instance, it is a logistic regression risk model that incorporates LDCT results and other clinical variables using NLST data. It notably improves discrimination power compared over models without screening results [6]. However, previous studies often focus on population level effects measuring average differences between groups (i.e., the average treatment effect [ATE]) of those with or without LDCT. In reality, the effect of the intervention would be different for each individual. Therefore, estimating the personalized individualized treatment effect (ITE) is crucial for tailored clinical decisions.

This study aims to build a counterfactual prediction model for lung cancer risk and mortality and to examine whether LCS as a hypothetical intervention reduces lung cancer risk and subsequent mortality. Specifically, a counterfactual model can elucidate the causal effect of an intervention (i.e., lung cancer screening in our context) on the outcome of interest (i.e., lung cancer risk and mortality) and answer the "what if" question: what if the patient goes through (or not) lung cancer screening, how would it impact the odds of the individual's mortality due to lung cancer?

2. Methods

2.1. Data source and study population

We obtained individual-level patient data from the OneFlorida+ Clinical Research Consortium [7], which contains robust longitudinal and linked patient-level electronic health record (EHR) data of ~16.8 million Floridians.

We consider adult patients (>18 years) who had done LDCT as the cohort of the interest and a control group of patients that are matched on age, sex and index date. Figure 1 shows the detailed study population selection flow along with a patient timeline of the covariates extraction period and the outcome observation window.

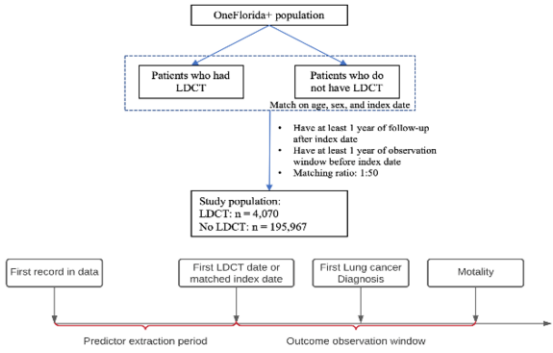

Figure 1. Study population selection flow chart and patient timeline of observation window.

2.2. Outcomes and covariates

We consider the first lung cancer diagnosis after the index date as the primary outcome,

and all-cause mortality the second. Covariates were obtained from a literature search on lung cancer risk factors [8]. These include age, sex, race/ethnicity, rural/urban status, smoking status, body mass index (BMI), Charlson comorbidity index (CCI), histories of hypertension, diabetes, heart disease, and stroke, and a series of pulmonary conditions, including cough, hemoptysis, chest pain, dyspnea, pleural effusion, lower respiratory tract infections, upper respiratory tract infections, chest infection, voice hoarseness, asthma, pneumonia, bronchitis, hay fever, emphysema, sputum cytological atypia, chronic obstructive pulmonary disease. Clinical conditions were defined based on ICD codes.

2.3. Causal modeling

We used 4 causal modeling approaches based on the potential outcomes framework for estimating ATE/ITE that address the selection bias differently.

Causal Forest (CF) [9]: Our initial method uses the propensity score for handling selection bias. CF, as a tree-based ensemble method, handles imbalanced and smaller datasets well. Each tree j of J trees in CF estimates treatment effects $e_j(\boldsymbol{x})$ by partitioning them to propensity score adjusted leaves based on their treatments. The CF estimates the treatment effects by averaging from J trees.

Representation learning-based models: We also use representation learning-based models, namely, Treatment-Agnostic Representation Network (TARNet), and Counterfactual Regression (CFRNet) [10]. These models learn a "*balanced*" representation that induces similar distributions between the treated and control groups. TARNet learns a shared covariates space for all instances and separate subspaces for instances of different treatment groups. CFRNet further applies the Integral Probability Metrics (IMP) as the distance between the distributions of treated and control groups and used to adjust the covariates subspaces.

Generative Adversarial Nets for ITE (GANITE): GANITE utilizes Generative Adversarial Nets to directly model the counterfactual distributions in [11]. This approach generates the proxy counterfactuals $\tilde{\mathbf{y}}_{cf}$ as the augmentation of the observational dataset D and uses the augmented dataset $\tilde{D}$ to optimize an ITE inference network **I** for the final estimation.

3. Results

We identified 4,070 who had LDCT and 195,967 matched controls from OneFlorida+, where patients had LDCT have a higher rate of lung cancer (3.9% vs. 0.5%).

Table 1. Baseline characteristics of the study population.

	Had no LDCT N=195,967	**Had LDCT N=4,070**	**Overall N=200,037**
Lung cancer	1,008 (0.5%)	157 (3.9%)	1,165 (0.6%)
Age	63.8 (6.31)	63.8 (6.25)	63.8 (6.31)
Female	97,839 (49.9%)	2,019 (49.6%)	99,858 (49.9%)
Hispanic	26,293 (13.4%)	154 (3.8%)	26,447 (13.2%)
non-Hispanic Black	33,740 (17.2%)	929 (22.8%)	34,669 (17.3%)
non-Hispanic White	101,391 (51.7%)	25,99 (63.9%)	103,990 (52.0%)
Other	34,543 (17.6%)	388 (9.5%)	34,931 (17.5%)
BMI	29.7 (6.86)	29.0 (6.94)	29.7 (6.86)
BMI unknown	53,751 (27.4%)	299 (7.3%)	54,050 (27.0%)

Current smoker	15,169 (7.7%)	1,886 (46.3%)	17,055 (8.5%)
Former smoker	29,220 (14.9%)	1,521 (37.4%)	30,741 (15.4%)
Never smoker	47,526 (24.3%)	48 (1.2%)	47,574 (23.8%)
Smoking status unknown	104,052 (53.1%)	615 (15.1%)	104,667 (52.3%)
Charlson comorbidity	2.40 (2.93)	3.58 (3.19)	2.42 (2.94)
Chronic pulmonary disease	48,340 (24.7%)	2,296 (56.4%)	50,636 (25.3%)

We first evaluated models' prediction performance with mean squared error (MSE). As shown in Table 2, CF has the largest (worst) MSEs on both tasks. The 2 representation learning-based methods, i.e., TARNet and CFRnet, achieved comparative performance, while GANITE outperforms all other methods. We investigated models' ATEs and find a consistent trend that the treatment effects of LDCT on lung cancer diagnosis are more significant than on mortality.

Table 2. MSE, ATE and IF-PEHE on predictions of lung cancer diagnosis and mortality.

Methods	Outcomes	MSE	ATE	IF-PEHE
CF	Lung cancer	0.027 (0.001)	0.033 (0.001)	1,310.76 (58.94)
	Mortality	0.029 (0.001)	0.022 (0.007)	552.23 (45.90)
GANITE	Lung cancer	0.003 (0.002)	0.144 (0.043)	795.94 (17.37)
	Mortality	0.005 (0.003)	0.063 (0.027)	501.66 (20.31)
TARNet	Lung cancer	0.006 (0.001)	0.251 (0.049)	872.36 (13.26)
	Mortality	0.007 (0.001)	-0.064 (0.053)	626.23 (17.25)
CFRnet	Lung cancer	0.006 (0.001)	0.526 (0.050)	841.21 (12.62)
	Mortality	0.008 (0.001)	0.119 (0.023)	597.24 (14.29)

4. Discussion

We evaluated 4 machine learning-based causal models for counterfactual prediction to address whether LDCT LCS reduces lung cancer risk and mortality, and then provide personalize estimation of the ITE for clinical decision support. Comparing the 4 models, DL-based models showed superior predictive performance compared to the traditional causal forest model. Among the DL methods, GANITE outperformed TARNet and CFRnet, providing more accurate ITE estimates based on the IF-PEHE metric.

Using OneFlorida+ network data, we confirmed that LDCT reduces lung cancer risk and mortality, aligning with RCT findings. This study is crucial as existing guidelines, largely based on age and smoking history, may inadequately stratify risk, leading to underutilization in high-risk patients and 'spill-over' in low-risk groups. Understanding an individual's ITE, precisely quantifying how LDCT can decrease lung cancer and mortality risks, empowers patients to make informed decisions tailored to their unique circumstances. With a causal model of ITE estimation, we can quantify the benefits of LDCT for each patient based on his/her unique medical characteristics. To showcase the models' advantages, we expanded beyond standard LCS age groups to include (1) age between 55 and 77, and (2) age below 55, coupled with different smoking status: (1) current smokers, (2) former smokers, (3) non-smokers, and (4) unknown status. As shown in Table 3, we present 6 patients as examples. Among these patients, we can see the patient "*6182*" and "*3471*" benefit most from having an LDCT by reducing more than 10% of their absolute probabilities of getting lung cancer. While even though patient "*853*" is in the same age group, but as a non-smoker, has almost no benefit gain from LDCT. On the contrast, patient "*105296*" although does not meet the age criterion for LCS, however, because of he is a current smoker with a history of asthma, bronchitis, and chronic obstructive pulmonary disease among other clinical indicators of high lung cancer risk, our model deemed him will benefit from LCS with a 7.3% risk reduction.

By using GANITE to estimate ITEs for each patient as described above, physicians can quantify how much a patient benefits from receiving an LDCT, thus making personalized treatment recommendations based on these quantified scores.

Table 3. Estimated ITEs of LDCT on lung cancer risk for selected patient samples.

Patient ID	Age group	Age	Smoking Status	$P_{(\mathbf{Y_0})}$	$P_{(\mathbf{Y_1})}$	$e(x)$
6182	Age>55 & <=77	66	Current smoker	12.7%	1.6%	-11.2%
3471		61	Former smoker	12%	1.5%	-10.5%
853		63	Non-smoker	<0.1%	<0.1%	<0.1%
400		70	Unknown	<0.1%	<0.1%	<0.1%
105296	Age <= 55	51	Current smoker	8%	0.8%	-7.3%
73000		54	Non-smoker	0.6%	<0.1%	-0.6%

$P_{(\mathbf{Y_1})}$: the probability of getting lung cancer if receiving an LDCT;
$P_{(\mathbf{Y_0})}$: the probability of getting lung cancer if NOT receiving an LDCT;
$e(x) = P_{(\mathbf{Y_1})} - P_{(\mathbf{Y_0})}$: benefits gained from receiving an LDCT versus NOT in terms of lung cancer risk.

5. Conclusions

In this study, we used causal inference methods to evaluate the impact of LDCT on lung cancer risk and mortality within the OneFlorida+ Clinical Research Consortium. We demonstrated the efficacy of these methods in reducing observational bias and provided estimates for ATE and ITE for a broader population. Future work will focus on model explainability to help patients and their providers understand what factors leads to the recommendations.

References

[1] Lung Cancer Statistics [Internet]. [cited 2022 May 9]. Available from: https://www.cancer.org/cancer/lung-cancer/about/key-statistics.html.
[2] Lung Cancer Fact Sheet [Internet]. [cited 2022 May 9]. Available from: https://www.lung.org/lung-health-diseases/lung-disease-lookup/lung-cancer/resource-library/lung-cancer-fact-sheet.
[3] Moyer VA, U.S. Preventive Services Task Force. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2014;160:330–338.
[4] US Preventive Services Task Force, Krist AH, Davidson KW, et al. Screening for lung cancer: US Preventive Services Task Force recommendation statement. JAMA. 2021;325:962–970.
[5] Huo J, Hong Y-R, Bian J, et al. Low Rates of Patient-Reported Physician-Patient Discussion about Lung Cancer Screening among Current Smokers: Data from Health Information National Trends Survey. Cancer Epidemiol Biomarkers Prev. 2019;28:963–973.
[6] Tammemägi MC, Ten Haaf K, Toumazis I, et al. Development and Validation of a Multivariable Lung Cancer Risk Prediction Model That Includes Low-Dose Computed Tomography Screening Results: A Secondary Analysis of Data From the National Lung Screening Trial. JAMA Netw Open. 2019;2:e190204.
[7] Hogan WR, Shenkman EA, Robinson T, et al. The OneFlorida Data Trust: a centralized, translational research data infrastructure of statewide scope. J Am Med Inform Assoc. 2021;ocab221.
[8] de Groot PM, Wu CC, Carter BW, et al. The epidemiology of lung cancer. Transl Lung Cancer Res. 2018;7:220–233.
[9] Wager S, Athey S. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. J Am Stat Assoc. 2018;113:1228–1242.
[10] Shalit U, Johansson FD. Estimating individual treatment effect: generalization bounds and algorithms. International conference on machine learning. 2017;3076–3085.
[11] Yoon J, Jordon J, van der Schaar M. GANITE: Estimation of Individualized Treatment Effects using Generative Adversarial Nets. International Conference on Learning Representations. 2018;