BRIEF

NOVEMBER 2021

ScreenNJ lung cancer screening: An innovative approach to identifying individuals eligible for low-dose CT scans

Krystin Sinclair

Aaron Truchil

Mouy Eng Kong Pan

Camden Coalition
of Healthcare Providers

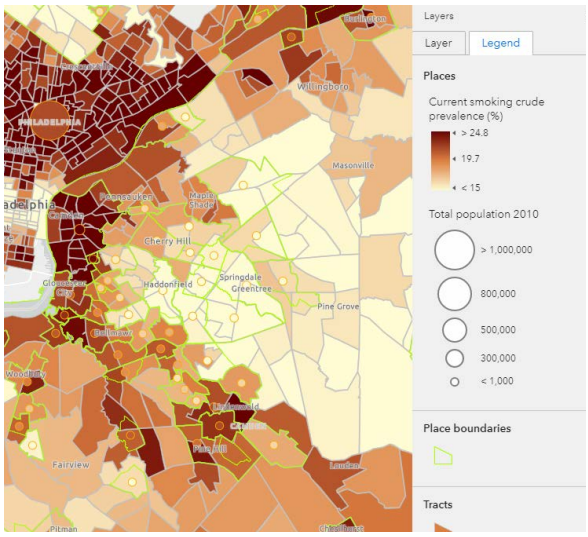
ScreenNJ lung cancer screening: An innovative approach to identifying individuals eligible for low-dose CT scans

Since 2019, the Camden Coalition has partnered with the New Jersey Department of Health (NJDOH) on its **ScreenNJ** initiative: a collaborative project aimed at reducing cancer incidence and mortality in New Jersey through expanding prevention and screening efforts. ScreenNJ focuses on colorectal cancer and lung cancer — both are among the most prevalent cancers in the state, and both have been shown to have better outcomes if detected early. Early on in our partnership, we encountered issues with how a patient's smoking history is documented, which made it difficult to determine who would be eligible for screening. We solved this challenge using natural language processing, a creative and innovative approach to creating structured data. We now have a process that proactively identifies individuals who would benefit from screening, which enables us to perform more targeted outreach for those most at-risk.

Introduction

Lung cancer is a leading cause of death from cancer for New Jersey residents. According to ScreenNJ, 9 out of 10 people with lung cancer **die from the disease** (NJDOH, 2018). As documented by the **New Jersey State Department of Health** (NJDOH, 2018), both Camden City and the surrounding county **experience significant disparities** (NJDOH, 2018) when compared to the state as a whole:

- Report smoking cigarettes (city, county, state)
    - 25%, 19%, 14%
- **CDC Places Project** (CDC, 2020) demonstrates the stark disparity in current smoking prevalence for Camden City versus its surrounding municipalities (see map)
- Lung cancer incidence (County, State)
    - 80 per 100,000, 64 per 100,000
- Deaths from lung cancer (County, State)
    - 40 per 100,000, 30 people per 100,000



As a result of these disparities, Camden City is a critical area for investing in more robust screening and prevention efforts.

The problem: Unstructured smoking data

Colorectal cancer screening is recommended for adults over 45, but lung cancer screening — which utilizes a low-dose CT scan (LDCT) — is not recommended for all adults. Per the **U.S. Preventive Services Task Force's** March 2021 recommendation, individuals must be between the ages of 50 and 80, and have significant histories of smoking (≥20 "pack years," which translates to smoking one pack a day for 20 years) in order to be eligible for lung cancer screening (USPSTF, 2021). "Pack years" is the common way that smoking history is recorded and is calculated by multiplying the number of packs smoked per day with the number of years smoked. While the criteria for identifying those eligible for screening is clear, accessing the necessary data is challenging.

In order to segment the population within those ages into those eligible for screening, providers must ask patients about their smoking history in a way that enables pack years to be calculated. Further, the data must be captured in a structured way within the electronic health records (EHRs). Free-text or unstructured forms make it exceedingly difficult to segment the population. Through the project, we found that structured "pack year" history data did not exist in the various data feeds contributing to the **Camden Coalition Health Information Exchange** (HIE). This information was only present as free text in patients' reports making it cumbersome to identify those eligible for screening.

The difficulty of accessing this data is not unique to us here in Camden; these issues occur throughout the country. Identifying smoking history through electronic health records (EHRs) requires **data mining** (Onega et al, 2017), as EHR smoking information is inadequately captured. As a result, physicians may be missing opportunities to discuss smoking cessation and lung cancer screenings with patients.

In **another study** (Golden, 2016) provider compliance was evaluated with documentation of smoking history and lung cancer referrals through a retrospective chart review based on EHR data from large, rural, community, and family practice clinics. This study found that roughly half of the patients' EHRs contained "pack year" information and that no patients who were deemed eligible for LDCT were given a referral for screening. The lack of comprehensive documentation was an issue. As Golden concluded, "Without completion of patients' smoking history, including length of smoking and packs per day, proper risk assessment for lung cancer cannot be completed."

Our solution: Natural language processing

While smoking history is a field that exists in the EHR system, it is not always filled out and not all smokers receive a smoking diagnosis code. Thus, relying on these fields will not provide the most accurate information regarding the amount of smoking that the patients do. In order for the Camden Coalition to identify patients eligible for lung cancer screening in our region, we endeavored to test the potential for natural language processing to generate structured data from the unstructured (free-text) data.

To do so, we:

- extracted two years worth of reports from the Camden Coalition Health Information Exchange (HIE), then
- identified report types (e.g., discharge summaries, radiology) that contain smoking history.

From this subset of over 5,000 patients taken from more than 7,000 encounters, we:

- built an extraction function to narrow down to the relevant text in the report,
- constructed an algorithm to identify the "pack year" number, and
- generated a structured data field.

We ran the algorithm across all of the reports and completed a random sample validation. We then compared the output field to the free-text report to confirm its validity. All in all, we identified 540 individuals (~10% of the patient population) who would be considered eligible for lung cancer screening.

We were later able to replicate this process for the **Trenton Health Team** (THT), who manages the Trenton HIE and uses the same HIE vendor (**CareEvolution**) as we do in Camden. THT also does cancer screening outreach work and was interested in finding more proactive ways to identify and engage eligible individuals. Working with THT and CareEvolution, we developed a similar extract of all reports from the Trenton HIE and employed the same steps with data. As a result, THT identified almost 4,000 patients with a high "pack year" history from the 40,000 HIE reports that contained the term "pack year." It is evident that the natural language processing approach has widespread applications.

Next steps

Both the Camden Coalition and our counterparts in Trenton now have the opportunity to reach out to specific individuals who have been flagged as eligible for lung cancer screening. We can do this through mailers, phone calls, or as part of other already established visits and interactions with patients. In the coming months, we hope to use the data we have access to as one of New Jersey's four **Regional Health Hubs**, including Medicaid claims data, to see if rates of LDCT scans — which have historically been low — tick upwards. We will then be able to connect the screening and outreach effort data to claims data to further evaluate whether our outreach efforts yield increases in screening.

References

Centers for Disease Control and Prevention. PLACES: Local Data for Better Health. **https://www.cdc.gov/places/index.html.** Published December 2020. Accessed October 2021.

Golden L. Documentation of Smoking History and Adherence to the USPSTF Recommendation for Lung Cancer Screening: A Retrospective Chart Review. 2016; **https://uknowledge.uky.edu/cgi/viewcontent.cgi?article=1130&context=dnp_etds**

New Jersey State Department of Health. Camden County Public Health Profile Report. **https://www-doh.state.nj.us/dohshad/community/highlight/profile/CigSmokAdlt.county/GeoCnty/4.html.** Published August 2018. Accessed October 2021.

Onega T, Nutter EL, Sargent J, Doherty JA, Hassanpour S. Identifying Patient Smoking History for Cessation and Lung Cancer Screening through Mining Electronic Health Records. 2017; DOI **https://doi.org/10.1158/1055-9965.EPI-17-0032.**

US Preventative Services Task Force. Final Recommendation Statement: Lung Cancer Screening. **https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/lung-cancer-screening.** Published March 2021. Accessed October 2021.

ScreenNJ. **https://screennj.org/.** Accessed October 2021.

About the Camden Coalition

We are a multidisciplinary nonprofit working to improve care for people with complex health and social needs in Camden, NJ, and across the country. The Camden Coalition works to advance the field of complex care by implementing person-centered programs and piloting new models that address chronic illness and social barriers to health and wellbeing. Supported by a robust data infrastructure, cross-sector convening, and shared learning, our community-based programs deliver better care to the most vulnerable individuals in Camden and regionally.

Through our National Center for Complex Health and Social Needs (National Center), the Camden Coalition works to build the field of complex care by inspiring people to join the complex care community, connecting complex care practitioners with each other, and supporting the field with tools and resources that move the field of complex care forward. The National Center's founding sponsors are the Atlantic Philanthropies, the Robert Wood Johnson Foundation, and AARP.

Camden Coalition
of Healthcare Providers

800 Cooper St., 7th Floor

Camden, NJ 08102

P 856-365-9510

F 856-365-9520