Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

doi:10.3233/SHTI220567

Identifying Determinants of Disparities in Lung Cancer Survival Rates from Electronic Health Record Data

Wanting CUI[1] and Joseph FINKELSTEIN
Icahn School of Medicine at Mount Sinai, New York, NY, USA

Abstract. The goal of this pilot study was to identify significant factors that affect disparities in lung cancer survival. A de-identified dataset was generated by querying electronic health records (EHR) from an academic medical center in New York City between January 2003 and November 2020. Socio-demographic characteristics, cancer stage, and genetic profile were analyzed using logistic regression. Two subsets of adult patients were identified: patients who were deceased less than 1 year after diagnosis and patients who survived over 5 years after diagnosis. Male, Black and Hispanic patients and those who were diagnosed in later stages were the people most susceptible to a shorter length of survival after cancer diagnoses. In addition, we identified three genetic oncodrivers (KRAS, EGFR and TP53) which were highly correlated with the length of survival after lung cancer diagnoses and their distribution was associated with race. We concluded that EHR data provide important insights on cancer survival disparities.

Keywords. Lung cancer, disparity, electronic health records, gene

1. Introduction

Lung cancer is one of the most common cancer in the United States for both males and females. It has a low five-year survival rate, compared to other common cancer types. Incidence and mortality rates are higher among Blacks as compared to Whites with lung cancer [1]. The major drivers of disparity in lung cancer survival include screening adherence, access to care and hereditary factors [2]. The goal of this pilot study was to identify major drivers that affect disparities in lung cancer survival using data from electronic health records (EHR).

2. Methods

A de-identified dataset was generated from Epic EHR system at the Mount Sinai Health System in New York City. We identified all adult lung cancer patients who were diagnosed between January 2003 and November 2020 within this dataset. We further identified 2 subsets of patients: patients who were deceased less than 1 year after diagnosis and patients who survived over 5 years after diagnosis. Logistic regression was performed to investigate the effect of demographic and cancer factors on patients'

[1] Corresponding Author, Wanting Cui, E-mail: wanting.cui@mssm.edu

duration of survival after cancer diagnosis. The independent variables comprised age, sex, race, cancer stage and genetic testing. The dependent variable was defined as a short-term survival. Somatic mutation data were used to identify genetic variants that affect lung cancer survival by analyzing variant distribution stratified by race and stage. We identified patients who had performed genetic testing and had test results in EHR.

3. Results

The analytical dataset contained 1099 patients: 750 of them survived over 5 years and 349 patients deceased within a year. In logistic regression, gender, race and cancer stage were important factors. However, age group and whether a patient has done genetic testing were not significant factors. Patients who were diagnosed at early stages of lung cancer were more likely to survive over 5 years than those who were diagnosed at advanced stages of lung cancer. In addition, Black patients and male patients had higher odds of being deceased in a shorter period of time after diagnoses, despite adjusting for age and cancer stage factors, compared to their White and female counterparts.

In genetic testing data, 214 (19.47%) patients had somatic genetic testing results. KRAS was the most common mutation. Around 47% of patients who underwent genetic testing had the KRAS gene mutation. In addition, 40% of patients who survived longer time had EGFR gene mutation, comparing to 12% of patients who survived a shorter time. Furthermore, patients who survived a shorter time have a higher proportion of TP53 gene mutations. There was higher proportion of Black patients (31%) than White patients (15%) with TP53 gene mutations who survived a shorter time.

4. Discussion

Male, Black and Hispanic patients who were diagnosed in later cancer stages were the people most susceptible to shorter length of survival after cancer diagnosis. KRAS was the most common genetic mutation among lung cancer patients and patients with TP53 mutations were at higher odds of being deceased in less than a year after cancer diagnoses.

5. Conclusions

Gender, race, cancer stage and somatic mutations were important factors that affects the length of survival of lung cancer patients after diagnoses. In addition, we found that Black patients had a higher proportion of the TP53 mutations, which was a gene mutation associated with short-term survival. Thus, future studies accounting for somatic mutations and availability of targeted treatment are warranted.

Reference

[1] Coughlin SS, Matthews-Juarez P, Juarez PD, Melton CE, King M. Opportunities to address lung cancer disparities among African Americans. Cancer Med. 2014 Dec;3(6):1467-76.
[2] Japuntich SJ, Krieger NH, Salvas AL, Carey MP. Racial Disparities in Lung Cancer Screening: An Exploratory Investigation. J Natl Med Assoc. 2018 Oct;110(5):424-427.