

FOR OFFICE USE ONLY

DATE RECEIVED: **02/18/2025**

DATE APPROVED: **02/18/2025**

FILE NUMBER: **#25-027**

APPLICATION FOR STATISTICAL CONSULTING

LAST NAME: **Kiel**

FIRST NAME: **Patrick**

DEPARTMENT (full name): **PHPR**

CAMPUS MAILING ADDRESS: **Heine Pharmacy**

PHONE: **3176457038**

EMAIL ADDRESS: **pkiel@purdue.edu**

YOUR PRIMARY POSITION AT PURDUE: **PhD Student**

Other:

(if a student) MAJOR PROFESSOR LAST NAME: **Preston** FIRST NAME: **Michael**

PHONE NUMBER: **7654945566**

MAJOR PROFESSOR CAMPUS ADDRESS (BLDG & DEPT): **PHPR / Heine Pharmacy**

MAJOR PROFESSOR EMAIL: **mapreston@purdue.edu**

HOW DID YOU FIND US: **Recommendation of my advisor or committee member**

LIST STATISTICS COURSES TAKEN AND STATISTICAL COMPUTING EXPERIENCE: **SPSS, R**

STAGE OF RESEARCH: **Analysis (all data have been collected)**

IF DESIGN STAGE IS COMPLETE, WAS A STATISTICIAN CONSULTED FOR DESIGN? **No**

PREVIOUS CONSULTANT – INSTITUTION/DEPARTMENT:

ESTIMATED NUMBER OF CONSULTING HOURS NEEDED THIS SEMESTER: **<5 hours**

EXPECTED COMPLETION DATE OF PROJECT: **5/1/2025**

IMPORTANT DEADLINE OR DUE DATES RELATED TO YOUR PROJECT:

THE RESULTS OF THIS RESEARCH WILL PROBABLY BE PUBLISHED AS:

Ph.D Dissertation, Journal Article

IS THIS RESEARCH SUPPORTED BY A GRANT OR CONTRACT? **No**

If so, give grant/contract title:

GIVE A BRIEF DESCRIPTION OF YOUR RESEARCH INCLUDING:

PURPOSE:

Research Questions and Hypotheses

1. Research Question 1: Does drug targeted therapy improve time-to-treatment discontinuation (TTD) and overall survival (OS) in patients with advanced NSCLC compared to carboplatin-based chemotherapy?
 - o Hypothesis 1: Access to genomically directed targeted therapy will increase time-to-treatment discontinuation and overall survival compared to carboplatin-based therapy in first line advanced NSCLC treatment.
2. Research Question 2: How do social determinants of health (SDoH) influence receipt of biomarker testing and the receipt of targeted therapies in patients with advanced NSCLC?
 - o Hypothesis 2: Patients with higher socioeconomic status (e.g., college education, higher income, or employment) have more access to biomarker testing and targeted therapies compared to those receiving conventional carboplatin-based chemotherapy. Conversely, disadvantaged SDoH factors, such as low income, unemployment, or disability, may reduce access to biomarker testing and targeted therapies.

DESCRIPTION OF VARIABLES TO BE MEASURED:

This retrospective cohort study utilizes data from the NIH All of Us Curated Data Repository Version 6, encompassing biomarker testing, clinical outcomes of TTD and OS, and SDoH factors from Jan 2017 to June 2022.

Eligible patients were aged > 18 years, diagnosis consistent with advanced NSCLC as the primary cancer, received first line systemic therapy with carboplatin plus pemetrexed or paclitaxel with or without pembrolizumab as the control group, or targeted therapy as generally recommended by the NCCN guidelines (Supplement). Data was extracted on advanced NSCLC diagnosis by including conditions from the standardized vocabulary: Primary malignant neoplasm of right lung, Primary malignant neoplasm of lung, Primary malignant neoplasm of left lung, non-small cell lung cancer, Metastatic non-small cell lung cancer, Malignant tumor of lung, and Squamous non-small cell lung cancer. Subclassifications were cross referenced with advanced metastatic therapeutic treatment and reviewed for accuracy.

Subjects were excluded if more than one cancer primary was documented, if treatment dates occurred before the date of primary diagnosis, or a diagnosis code or treatment was not associated with squamous or adenocarcinoma cell carcinoma.

Study Measures

Time-to-Treatment Discontinuation and Overall Survival

TTD was measured as the number of days from initiation of first line therapy to the last documented date of administration. Censoring occurred for patients who remained on treatment within 120 days of July 1, 2022. OS was measured as the number of days from initiation of first line therapy to the documented date of death. Patients without a death event were censored at the end of the database or last observation.

Baseline Demographics and Social Determinates of Health

Variables of interest included race, ethnicity, smoking status, disability, education level, employment status, marital status, neighborhood safety, insurance type, and income. Patient race variable was categorized into white vs. not white. Ethnicity was classified as Hispanic and non-Hispanic. All other survey data was self-reported. Disability included at least one of the following categories, deaf, blind, difficulty with errands alone, difficulty concentrating, difficulty dressing or bathing, difficulty walking or climbing stairs. Education level was categorized as any college education or not. Employment status was defined as employed, unemployed, or retired. Sex was documented according to birth. Marital status was self-reported and classified as married or not. Neighborhood safety categorized as safe or not. Self-reported income was categorized as > 75K or less than 75K.

Baseline characteristics used descriptive statistics and were compared using Student t test or chi-square test as appropriate. Missing data were not imputed and not included in statistical testing. Statistical significance was defined by a 2-sided alpha level of 0.05 or by using a 2-sided 95% confidence interval. TTD and OS comparisons were made between patients who received carboplatin-based chemotherapy and a targeted therapy agent and described by using Kaplan-Meier analysis. Log-rank tests were conducted to perform unadjusted comparison between groups. A Cox proportional hazards regression model was fitted to estimate hazard ratios (HRs) and 95% confidence intervals (CIs). The following covariates were planned for the model: targeted therapy (yes vs. no), age group (>60 vs. ≤60 years), race (White vs. non-White), ethnicity (Hispanic/Latino vs. not Hispanic/Latino), sex at birth (Male vs. Female), smoking status (current/former smoker vs. never smoker), and biomarker testing (yes vs. no). Social determinates of health were evaluated with multivariate logistic regression to estimate the odds ratios (ORs) and 95% CIs for the association between SDoH variables and the receipt of biomarker testing and the receipt of a targeted therapy agent.

RESEARCH QUESTIONS THAT YOU WANT TO ADDRESS USING STATISTICAL METHODS:

Two major areas of assistance:

1. Guidance on additional analysis on missing data.
2. Best approach of either backward, forward, or stepwise regression are variable selection methods used in multiple linear regression to choose the best subset of predictor variables.

STATISTICAL ISSUES:

Two major areas of assistance:

1. Guidance on additional analysis on missing data.

2. Best approach of either backward, forward, or stepwise regression are variable selection methods used in multiple linear regression to choose the best subset of predictor variables.

ADDITIONAL INFORMATION YOU THINK WOULD BE HELPFUL:

ATTACHMENTS:

[No Attachment](#)

