Peer review file
Article information: https://dx.doi.org/10.21037/jtd-21-804

Reviewer A

Comment 1: Positive rate would be different across the ethnicity. Caucasian has lower rate of EGFR positive status compared with Asian. Also, preceding studies have shown that rate of oncogenic driver screening test differs depending on the countries. Discussion on this point might cultivate the manuscript and will help the readers to interpret this study more appropriately.
Reply 1: Thank you for your feedback. The following text has been included in the background section of the manuscript.

"The prevalence of *EGFR* mutations in advanced NSCLC is variable depending on the geographical region and ethnicity of the patient. In patients with adenocarcinoma histology and of Asian ethnicity, prevalence can be as high as 50%, compared 15-20% in Caucasians. (14,15) The Canadian population is comprised of many different ethnicities and *EGFR* mutation rates occur in approximately 20.6% of non-squamous patients. (16) Testing for the *EGFR* mutation has evolved overtime in Canada. In Ontario, Canada testing for the *EGFR* mutation began in 2010, and reflex testing at the level of the pathologist for *EGFR* mutations in non-squamous NSCLC has been implemented between 2011-2014 in most centers. (17) In 2015, the prevalence of *EGFR* mutation testing was approximately 72% for advanced non-squamous NSCLC patients at one institution in Ontario, Canada."
Changes in the text: Line 133-142 have been added.

Comment 2: The cohorts contain squamous and large cell histology. They would almost have negative EGFR status. What if the authors perform the similar research limited to this population, if they cannot differentiate the EGFR status in the whole population?
Reply 2: We are interested in all NSCLC patients not restricted by pathological subtype as this is the population eligible for later-line Erlotinib, and although restricting by non-adenocarcinoma subtype may eliminate more unknown *EGFR* negatives, it will detract from the power and generalizability of the study.
Changes in text: No changes have been made.

Comment 3: The data on ED visit may be an important insight gained from this study. If available, please show more detailed descriptions on this point, not just as ED visit (the number of times, the reasons etc.)
Reply 3: We agree that this information would be very valuable and insightful. Unfortunately, the data we have access to does not contain information on the reasons for visits. We have added data on the frequency of ED visits in the cohort as below
"In our cohort patients had a mean number of ED visits of 1.9 (SD 2.25) with a range of 0-42 visits."
Changes in text: Line 318-319

Comment 4: Table 1: Please specify the ethnicity of "general population"
Reply 4: The specific ethnicity of the general population is unknown. This has been updated
Changes in text: See table 1

Reviewer B

Comment 1: The title does not reflect well the conclusion of the results. Therefore, it should be changed accordingly. In addition, as the study describes the results in a particular population of Canada where the health insurance still allows monotherapy in advanced adenocarcinoma despite the absence of information on EGFR mutation, the title must also reflect this (e.g., in Canada).
Reply 1: The title has been changed to "Second and later-line erlotinib use in non-small cell lung cancer: real world outcomes and practice patterns overtime in Canada"
Changes in text: Line 4

Comment 2: Why the "area of residence" was included as covariates in the analysis. Was performance status included as a covariate?
Reply 2: Ontario Canada is a large province with various health "areas of residence". Occasionally in other studies there have been differences between outcomes or treatments in patients depending on where they live so this co-variate was included to correct for any differences. No, performance status was not included as this was not available in the database.
Changes in text: None

Comment 3: It was not mentioned under the subtitle "Other covariates" whether disease stage (IIIB, IV; Table 1) was included as covariates?
Reply 3: No disease stage was not added as a covariate. We included only patients that received palliative chemotherapy protocols (not concurrent chemoradiotherapy patients), and incurable stage IIIB and stage IV were not separated from each other.
Changes in text: No changes

Comment 4: The authors should further discuss why, in the erlotinib cohort, the hazard ratio for death was high despite the overall survival being longer than in other groups. They discussed a little on page 11, but there is no cited reference to support their explanation.
Reply 4: Please see line 329 to line 342 of the discussion. Here we discuss how despite the crude difference in the OS, this is not attributable to erlotinib and is driven by other factors. We have added some more text for clarity.
Changes in text: See line 331-332
"To our knowledge, this is the largest retrospective study of real-world second or later-line erlotinib use in unselected advanced NSCLC patients. Despite the longer crude survival difference in patients treated with erlotinib, the hazard for death suggests this difference was not attributable to erlotinib treatment, and rather the increased survival seen in the erlotinib group is reflective of the fact that these patients simply lived long enough to receive erlotinib. Review of the literature suggests marginal benefit of erlotinib therapy in the unselected or EGFR wild-type population. Many of those studies excluded patients who would typically be treated with erlotinib in the real world (eg. those with brain metastases, poor performance status, organ dysfunction, recent radiation). However, these real-world patients are included in our analysis, which may account for the differences seen. Patients treated with erlotinib in our study also had higher relative risk of visiting the emergency department. This likely reflects increased healthcare utilization by advanced cancer patients receiving ongoing active medical management at end of life, rather than erlotinib toxicity."

Comment 5: In figure 3, why did the authors not show the hazard of mortality over time in the non-erlotinib cohort?
Reply 5: Our primary objective was to examine the association between erlotinib and mortality. This was accomplished using the estimate of the hazard ratio (comparing those on erlotinib against those not on erlotinib), where erlotinib was appropriately treated as a time-varying exposure. A secondary objective was to better understand how the intensity of mortality varied over time once erlotinib therapy began. As a result, our decision to explore the hazard function for those on erlotinib was made a priori. Based on the estimated hazard ratio of 1.89, we would expect the estimated y-coordinates of the hazard function plot for those not on erlotinib to be approximately 1.89 lower than seen in Figure 3.
Changes in text: None

Comment 6: Also, is it possible to show the survival curves of each cohort?
Reply 6: The data in this study is retrospective and to do a survival analysis needs to be corrected for covariates as we have done. Once you add time varying covariates (chemotherapy history and erlotinib use), from a statistical perspective you can do longer report median overall survival or generate survival curves, you can only report the hazard ratio. See text line 296-298.
Changes in text: None