Reviewer's report

Title: Underutilization and disparities in access to EGFR testing among Medicare patients with lung cancer from 2010-2013

Version: 0 Date: 24 Jul 2017

Reviewer: Robert Hines

Reviewer's report:
Comments to authors: The authors sought to evaluate the prevalence and predictors of EGFR testing in lung cancer patients. This outcome is relevant as it predicts response to EGFR tyrosine kinase inhibitors (TKIs) which are a newer class of molecularly targeting agents distinct from systemic chemotherapy. However, only specific patients will benefit from EGFR TKIs and even among patients who initially respond, "most patients inevitably progress despite initial dramatic and rapid response to EGFR TKIs" (Lee DH. Treatments for EGFR-mutant non-small cell lung cancer (NSCLC): The road to a success, paved with failures. Pharmacol Ther. 2017;174: 1-21.).

The current recommendation from the National Comprehensive Cancer Network only recommends EGFR testing in metastatic lung cancer patients with specific recommendations based on histologic subtypes with further consideration based on a combination of histologic subtype and behavioral exposure factors (squamous cell carcinoma, in never smokers or small biopsy specimens or mixed histology). In a 2013 article, another guideline-issuing group was less restrictive regarding tumor stage (stating that testing was "encouraged" in patients with Stage I-III disease) but the recommendation was limited to tumors with an adenocarcinoma component or where an adenocarcinoma component could not be ruled out (Lindeman NI, Cagle PT, Beasley MB, Chitale DA, Dacic S, Giaccone G, Jenkins RB, Kwiatkowski DJ, Saldivar JS, Squire J, Thunnissen E, Ladanyi M. Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. J Thorac Oncol. 2013;8: 823-859). However, the authors had no ability to subset the data based on these recommendations given the limitations of the dataset. There were also other limitations that compromise the ability to make valid inferences from study results.
Some study limitations could be overcome by using the SEER-Medicare dataset. The authors state that they did not use this dataset because it represents only 28% of the lung cancer population. However, internal validity is a prerequisite for establishing external validity. This study generates serious concerns regarding internal validity making external validity a moot point.

Abstract:
Background: Even though the abbreviations is commonplace, I would define EGFR the first time they it used in the Abstract and manuscript. Less important to define KRAS as the name does not equate to what it actually is (i.e., a proto-oncogene).

Background: Here, you state that EGFR testing is indicated for all newly-diagnosed, metastatic lung cancer patients who are candidates…"

Grammatical error: "Patients who lived in the Boston", "the" should be deleted. MA can be used to abbreviate Massachusetts as state abbreviations are commonplace.

I prefer the term association or increased/decreased odds vs. the term likelihood to describe odds ratios. It is true that odds ratios and probability (risk) ratios both describe the likelihood of an outcome. However, most people equate likelihood to probability (although there is debate concerning this point among the biostatisticians). Odds ratios, that is the ratios of odds (the probability of something occurring divided by the probability of something not occurring) are often interpreted the same as risk ratios (probability of occurrence in the exposed divided by probability of occurrence in the unexposed), but they are not the same (they will, however, approximate each other for rare outcomes). Therefore, if you are referring to a probability, should not equate an odds ratio with an increased/decreased probability. You can state an increased/decreased odds. Although odds ratios are commonplace and still considered appropriate for an analysis of dichotomous outcomes, if you are truly wanting probability (which I would argue would be a better choice here), you should perform binomial regression. Then you can provide estimates of effect that have more intuitive meaning for the reader. Using binomial regression, you can validly refer to the ratio measure x times more/less likely to occur in one
group vs another. In summary, you can keep the odds ratios, but I would suggest you refer to increased odds or x times the odds rather than use the term likelihood which most readers will equate with probability. For more on this, see Sainani KL. Understanding odds ratios. PM R. 2011;3: 263-267. If you're wanting to learn more about binomial regression, see Woodward M. Epidemiology : study design and data analysis. Third edition. ed. Boca Raton: CRC Press,Taylor & Francis Group, 2014.

Background:
1st paragraph: "Lung cancer causes serious morbidity and mortality in 1.7% of Medicare beneficiaries." What does this sentence mean? There is no way to interpret this sentence because "serious morbidity" is not defined. Do you mean only 1.7% of patients diagnosed with lung cancer will die due to their disease?

1st paragraph: "Studying lung cancer molecular test utilization within Medicare claims provides a unique opportunity for a comprehensive, population-level analysis of precision medicine testing." OK, but there are limits to this recommendation. For example, "Use of the EGFR-TKIs gefitinib, erlotinib, and afatinib is limited to patients with adenocarcinomas who have known activating EGFR mutations" (http://emedicine.medscape.com/article/1689988-overview).

2nd paragraph: "Over the last decade, molecular testing of lung tumors has become an essential component of diagnosis and treatment of advanced non-small cell lung cancer (NSCLC)." A reference is needed to support this assertion. In addition, the nuances of testing associated with histologic subtype and disease stage is not cited.

4th paragraph: "From 2011 through 2013, EGFR testing was indicated for all patients with newly diagnosed metastatic adenocarcinoma of the lung being considered for first-line therapy with an EGFR tyrosine kinase inhibitor. This indication corresponded to approximately 35% of all new lung cancer cases." Ok, here you do get more specific regarding the guideline recommendation and the proportion of lung cancer cases this represents. However, a limitation of the dataset used for the study was the inability to subset tumors based on histologic subtype and disease stage.
5th paragraph: This paragraph is not clearly linked to the rest of the background section-I would delete it.

6th paragraph: "EGFR mutations are less frequent in squamous cell carcinomas therefore the guidelines suggest testing only those patients with squamous histology whose clinical or demographic characteristics (e.g., absence of smoking history, Asian descent) indicate an increased likelihood of mutations." Again, the ability to subset the data to directly test patients for whom the guideline is intended is not possible.

Last paragraph: You identify the objectives of the study in this paragraph. However, this does not include KRAS which was mentioned in the objectives of the abstract.

Last paragraph: The explicit evaluation of the second objective was not addressed.

METHODS

Data Sources
Even though 93% of lung cancer cases were incident cases in 2011-2013, I would exclude 2010 which includes both incident and prevalent cases to ensure an "apples to apples" comparison.

Variables
To ensure comparability with 2010-12, for 2013, I would use the same methodology used to identify combined EGFR/KRAS tests. That is, I would combine these two tests into one outcome as was done for the earlier time period.

RESULTS

Molecular Testing from 2010-2013
In Table 1a, I would label code "83912" as EGFR/KRAS because that is what it represents.
"There were 13,568 patients identified in 2011 and 14,302 patients identified in 2012 who underwent at least one molecular test." Ok, I now have finally figured out Table 1a and what the total column represents. I think you need to have clear, concise language explaining that the total column represents individual patients that received molecular testing (which corresponds to "83912" + EGFR + KRAS - Multiple).

2nd paragraph: Just to clarify, does the number given for "Claim for surgical pathology analysis" and "Claim for reporting a molecular test" correspond to incident cases of lung cancer diagnosed in that year or could those come from previously diagnosed cases? For example, for 2012, does the 13,818 claims for molecular testing out of the 155,408 claims for surgical pathology represents claims for those 227,929 diagnosed with lung cancer in that year? What about those patients who were diagnosed in late 2011 (for example) who could have had a claim in 2012?

2nd paragraph: "The absolute number of claims for molecular testing decreased slightly in 2013. This decrease may be explained by limitations in the data. Patients diagnosed in December of 2013 may have been tested in January 2014 or there may be an expansion of next generation sequencing, which would not be billed with the gene-specific billing codes." OK, does that limit our ability to compare this year to the previous years?

Patient characteristics by lung tissue analysis and molecular testing
1st paragraph: "Among the patients identified from 2011 through 2013, 642,570 (93%) had a prior year of claims data without any indication of a lung cancer diagnosis, which illustrates these patients represented mostly newly diagnosed cases." This has already been described in the METHODS.

2nd paragraph: National incidence data is reported for all cancers, therefore, I don't really see what this paragraph is adding.

3rd paragraph: "Identifying patient-level differences in access to surgical pathology is important because these patients will not have access to lung tissue molecular testing." Avoid interpreting results in the RESULTS section, save for DISCUSSION if you think it is worth mentioning.
4th paragraph: "There were small but statistically significant differences by age, race, Medicaid status, and risk score and testing." Change "and testing" to "according to testing status" or something of the like to differentiate it from the list of variables.

4th paragraph: "Beneficiaries under age 55 were the least likely to be tested (5.9%), which may be explained by an earlier stage of diagnosis or by these patients being diagnosed and tested prior to enrolling in Medicare." Are younger lung cancer patients more likely to be diagnosed with earlier disease stage? Also, even I would save your interpretations for the DISCUSSION.

4th paragraph: You do not need to use the text to point out every difference and multivariable results are more important than bivariable.

Regional differences in lung cancer molecular testing
I think it is useful to have a map, but I would combine all years and use the same methodology used in years 2011-2012 to have one map. It is a bit busy to have four maps and unclear as to what the reader is supposed to "take home" from these four images. In my opinion, there is way too much text devoted to these results.

Factors that predict utilization of lung cancer molecular testing
1st paragraph: Suggest that you change "likelihood" to "odds".

Table 4. What was the reference group in calculating the odds ratio? These numbers are meaningless without defining the reference group.

2nd paragraph: "Clinical procedures had the next strongest correlation with testing." Given that these procedures are not mutually exclusive, there could be collinearity? Is it possible to collapse categories resulting in mutually exclusive patient groups?

2nd paragraph: "For a 0.10 increase in risk score, there was a 2.6% decrease in likelihood to be tested," (which corresponds to 23% decreased odd of testing for a 1-unit increase in risk score). Make sure your text is consistent with your table.
3rd paragraph: "Medicaid recipients, who disproportionately include Black patients, had an OR of 0.74 (CI 0.72-0.77)." Was there an interaction between Medicaid status and race?

DISCUSSION
5th paragraph: "Clinical reasons may explain variation in testing by age and level of comorbidities. Beneficiaries who were enrolled in Medicare prior to age 65 have disabilities, including endstage renal disease, that may influence treatment decisions." Yes, but didn't you previously indicate (in the Results) that for many patients under 65, the disability associated with their lung cancer diagnosis is what made them eligible for Medicare?

8th paragraph: "However, claims data do not contain clinical information that is relevant to eligibility for EGFR testing, such as the date of diagnosis, cancer stage, histological subtype of disease, and tissue availability." This was a significant study limitation which raises serious concerns regarding the ability to obtain valid inferences from the data.

8th paragraph: "However, if we limited our analysis to Surveillance, Epidemiology, and End Results (SEER) linked Medicare data, it would have represented only 28% of the US lung cancer population. Our goal was to provide a population-level analysis." See previous comment: the primary concern of an epidemiologic study is to obtain valid estimates of association. The ability to generalize beyond the study population is a secondary concern of which internal validity is a prerequisite.

In the future, please provide a concluding paragraph to summarize the importance of these results and clearly state what the "take home message" is for your study.

Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.
No

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.
Yes
Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.
No

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.
I am able to assess the statistics

Quality of written English
Please indicate the quality of language in the manuscript:
Needs some language corrections before being published

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?
2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?
3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?
4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?
5. Do you have any other financial competing interests?
6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests.
I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal