Reviewer A

Comment #1: "Major issue with this study is that it lacks the specific cause of death AND that it calculates overall survival rather than 5-year survival."

Reply #1: We acknowledge that a limitation of this retrospective study is that the specific cause of death is not available in this database. We have addressed the issue of the unknown specific cause of death in the discussion section. We agree with the reviewer that 5-year survival is a valuable endpoint. To address this, we have revised our initial KM survival analysis to include 5-year overall survival analysis based on stage and other SDOH of interest (race, marital status, insurance status, and gender). We found overall 5-year survival was worse for Black compared to White individuals (17.3% vs. 29.2%, respectively); worse for males compared to females (23.9% vs. 31.9%); worse for those uninsured and Medicaid compared to private only insurance (15.2%, 21.8%, and 32.7%); and worse in single compared to married individuals (19.5% vs. 30.4%).

Changes in the text: Methods section and discussion section – 4.4 Strengths and Limitations, Line 537. Five year survival is addressed throughout the "Results" and section Figure1.

Comment #2: Also, unlike what authors claim SDOH is construct that include a lot more than just insurance status (which is not appropriately categorized in this study) or marital status or race.

Reply #2: We agree that SDOH encompasses the individual variables studied here (race, gender, insurance, and marital status) and many other potential variables. The individual variables suggested by the reviewer (income and education) were not documented in the IUSCCC database nor were they easily obtained through the EMR. Numerous studies provide a rationale for the use of insurance status as a surrogate for lack of/low educational attainment and low income. For instance, Healthy People 2020 notes that those aged 25-64 years of age with an advanced degree have the highest rate of health insurance coverage among educational attainment groups (*http://www.healthypeople.gov/2020/leading-health-indicators/2020-lhi-topics/Access-to-Health-Services/data*). Another study by Goldman et al observed the growing problem of non-access to health insurance was increasingly concentrated among the less educated (Goldman, D., Smith, James P. (2011). The Increasing Value of Education to Health. *Soc Sci Med. 2011 May ; 72(10): 1728–1737. doi:10.1016/j.socscimed.2011.02.047.*)

We decided not to include census track level variables, because we wanted to focus on known individual level SDOH as opposed to population level SDOH. However, we agree with the reviewer that these and similar measures can help to further define the impact of SDOH on patient care. Our ongoing research using neighborhood level SDOH to determine their impacts on lung cancer outcomes is currently being done.

Changes in the text: Edits to Discussion lines 453-462 and 526-528.

Comment #3- The first line of key findings claim that racial disparities are "mediated" by SES factors. This is not a mediation analysis and this is a wrong terminology to be used. This term needs to be changed to "explained". The same should be done regarding line 105. The term "mediated" is epidemiologically wrong in this context.

Reply #3: We agree that "mediated" is the wrong terminology to use. Changes are made in the text. Please see below.

Changes in the text: Terminology changed in the text: Highlight Box☐Key findings and in the text as indicated (line 125).

Comment #4- Lines 125: it should be specified if race and ethnicity are exclusive, Non-Hispanic White and Non-Hispanic Black?

Reply #4: Thank you for the comments. Table 1 includes the demographic characteristics of the cohort, and the ethnicity (Hispanic vs non-Hispanic) is shown. For both Black and White race, Hispanic ethnicity made up only 0.25% and 0.58%, respectively, and therefore, subsequent analysis was made using the broader categories or race, since the largest percentage of these two racial groups were non-Hispanic.

Changes in text: Further clarification of how ethnicity was addressed and is included in the Methods, lines

153-156.

Comment #5- The study says that they include Hispanic ethnicity but in none of the tables they report Hispanic findings. Given that only about 0.5% of the cohort are Hispanic, I suggest excluding those with Hispanic ethnicity.

Reply #5: You are correct. We showed ethnicity in Table 1 to describe the cohort demographics, but ethnicity was not analyzed as a separate variable given the small percentage included in both racial groups. We have addressed this in the methods section.

Changes in text: We changed this in the text and removed mention of ethnicity as a variable of interest as it was removed from subsequent analysis due to the small sample size. This was addressed in section 2.2.1, lines 153-156.

Comment #6- The authors need to provide citations and do some literature review to appropriately categorize insurance status. It is inappropriate to group Medicare and Medicaid together. Medicare covered individuals usually have poorer prognosis considering they are 65 plus and older. Medicaid covered individuals and Medicare/Medicaid dual eligible individual have some of the worst cancer outcomes. Medicaid people are usually younger than 65. Researchers suggest Medicaid associated negative cancer outcomes may be related to the fact that the coverage gauge poverty. Also, the two have two different payment schemes. Below this document, I am suggesting some articles that will help.

Reply #6: We agree with Reviewers A and C's comments about the typical classification of insurance status and rationale. Particularly, Hines et al. and Ko et al. classified insurance status as those with "Medicaid/ uninsured" vs "other forms" or "private". We also acknowledge that Medicaid status is a surrogate for individual poverty level, but may differ from uninsured. For instance, for Indiana residents, Medicaid covers oncology services, including cancer prevention, diagnosis, therapeutic treatment, rehabilitation, and palliative care. Upon reflection, we agree with many of the reviewers' concerns. Specifically, our initial classifications may introduce confounders which make evaluation of insurance as a SDOH on lung cancer treatment/management less impactful, particularly by grouping those with Medicare and Medicare into a "public insurance group" and not assessing these two types of insurance separately. We have therefore reclassified insurance to provide a more granular evaluation of insurance on patient outcomes. In this revised manuscript, insurance is categorized as: Medicaid only, Medicare only, Medicare and Medicaid (Medicare/ Medicaid), Self-pay/uninsured, and Private. We believe this better illustrates the impact of insurance on lung cancer outcomes in this cohort.

We completed subsequent analyses using these new insurance categories. For instance, our new analysis further shows increased diagnosis of advanced stage NSCLC (stages III and IV) in those uninsured, followed by those with Medicaid only and Medicare/Medicaid (71.43% vs. 78.24% vs. 21.76%). Survival analysis also showed the worse overall survival was among the uninsured individuals, followed by those with Medicaid/Medicare and Medicaid only.

Changes in text: Rationale for insurance classifications are described lines 158-161. Insurance status was recategorized in Table 1 to include the following categories: Medicaid only, Medicare only, Medicare/ Medicaid, Self-pay/uninsured, and Private. Subsequent analysis was redone for all prior analysis that used the prior insurance categories with edits made throughout the results as needed.

Comment #7- Lines 133-135: age categories seem arbitrary and irrelevant. They need to reclassify to less than 50, 50-64, 65 and above. Given that comorbidity is not available, this type of classification is better able to group people together. Also 65 is the age to be eligible for Medicare and is a better cutoff point.

Reply 5: We agree with the reviewer. We have made changes in Table 1, keeping the lower cut-off of age (<40) separate, given the younger age of Black lung cancer patients at the time of diagnosis. These ages are now characterized as <40, 40-49, 50-64, >/= 65.

Changes in text: Changes made in Table 1 and line 168-169 of the manuscript.

Comment #8: Overall mortality needs to change to all-cause mortality. This change should be implemented through the entire manuscript.

Reply #8: We have used overall survival instead of overall mortality or all-cause mortality as this survival is

the preferred outcome particularly for patients with cancer. The literature supports use of the terms overall survival and all-cause survival interchangeably and therefore, we did not make changes in the text.
Changes in text: None

Comment #9: -   Lines 159-183: the authors do not explain why they are calculating overall survival while 5-year survival is better benchmark for NSCLC (see some the citations below). This is particularly concerning given that they do not have lung-cancer specific death and also because only less than 13% of the data in their population is younger than 50 and the majority are older.
Reply #9: We agree that 5-year survival is important. Please refer to Reply #1 above.
Changes in the text: Please refer to "changes in the text #1". Results section Figure 1 and Supplemental Figure 2.

Comment #10: Lines 180-181: It needs to be specified if the proportional hazard assumption was tested for? If not, it needs to be tested for violation. If there is serious violation with regards to exposures of interest (stage, race, insurance, marital, smoking) , then the violation needs to be addressed.
Reply #10: Thank you for this question. We checked the proportional hazards assumption by plotting log{-log[S(t)]}vs log(t) for categorical variables and plotting of weighted Schoenfeld residuals versus time for continuous variables. If there was very strong evidence that the proportional hazards assumption was violated, then we included an interaction of the variable with time into the regression to model the dependence of the hazard ratio on time.
Changes in the text: No changes made as this was done in the prior analysis.

Comment #11- Most tables are very confusing. In all tables, the reference category should be clarified instead of for example saying "male vs female" similar to what has been done in supplemental table 9
Reply #11: It is not our intention to confuse the readers, so we appreciate this feedback. We have added a "reference" category to the appropriate tables.
Changes in the text: Results section. All tables now include reference values.

Comment #12- For the survival analysis, it is important to repeat the analysis stratified by stage. Stage is one of the most important drivers of surgery and survival and therefore it deserves a closer look through stratification.
Reply #12: We agree that stage is one of the most important contributors to survival and have stratified our KM curve based on stage. As expected, all-cause and 5-year survival was worse for those with advancing stages of NSCLC.
Changes in the text: Changes made in results section (lines 364-367), Supplemental Figure 2.

Comment #13- For the survival analysis, those who died within the first 2 months must be excluded as they will cause serious survival bias issue. Also, those who died within the first 30 days should be excluded as they may caught by the registry in autopsy.
Reply #13: Those who die within the first 2 months likely represent those with advanced stage disease. Therefore, additional survival analyses were stratified based on stage of diagnosis as in #12. Our registry did not include autopsy results. Survival analysis was similar when we included and excluded the 212 patients that died within the first two months. Since one of our outcomes of interest was the impact of SDOH on staging and overall survival, we felt it was important to include those who died within the first 2 months as these SDOH could be potential identifiers of those with worse overall survival.
Changes in the text: Not shown.

Comment #14- The study must acknowledge that not having comorbidity and cause of death as major limitations.
Reply #14: We agree and have included this as a limitation of our retrospective study.
Changes in the text: This was addressed in the Discussion section – 4.4 Strengths and limitations, Lines 532-536.

Comment #15- Citing only 26 studies for the well-studied topic of racial disparities suggest that authors have not done their due diligence in reviewing other studies. That is why they claim that there are few studies

that included similar variables. This statement is not accurate. I suggest to strengthen their introduction and discussion, to expand their literature review. Below are my suggestions to start with although there are more on this topic (some of my suggestions are to help you address other issues discussed above):

Reply #15: We appreciate the reviewer's point and have added additional references that support our findings in complementary studies examining the impact of neighborhood level and individual level SDOH on other cancers including bladder, breast, and lung (see below). We reframed our statement in the text to clarify that our study provides a comprehensive evaluation of these specific SDOH on reasons for not receiving recommended treatment for early stage NSCLC. The findings that specific SDOH impact receipt of guideline-recommend treatments will guide future prospective studies to directly explore reasons for the observed disparities in guideline concordant treatment of early-stage disease based on these SDOH.

Changes in the text: Discussion section, multiple edits and References section – we included multiple additional contemporary references as detailed in the marked version of the resubmission.

Reviewer B

Thank you for your important work showing that some of the inequalities are caused by health care. Manuscript was easy to read and language was fluent.

Comment #16: Table 1. Should include BMI, performance status, CCI-index, limited FEV1. If this data is not included to the registry, you should consider them as possible limitations. Definition of smoking status is unusual.

Reply #16: Thank you for your overall enthusiasm for the findings in this manuscript. We agree that these patient characteristics are important to lung cancer outcomes. These are not available in our database. We acknowledge this is a limitation of this retrospective study.

Changes in the text: Discussion section lines 542-546.

Comment #17: Table 4. Could you present how many were treated according to guidelines by race and stage n/row percentage to understand your data better. CI 95% are wide and all data in the table might have only marginal interest for readers outside U.S.

Reply #17: Thank you for this suggestion. We have made these changes to Table 4 and believe it makes it easier to interpret some this data.

Changes in the text: Refer to results section table 4.

Comment #18: Table 5 and 6 could you please consider including statement which variables were included in multivariate analysis

Reply #18: The variables in tables 5 and 6 were those included in the multivariable logistic regression analysis, which are based on those variables that showed statistical significance by univariable analysis shown in the supplemental tables. We have made changes below Tables 5 and 6 to reflect the statistical methods.

Changes in the text: Statement made under the referenced tables. See changes to the text below Tables 5 and 6.

Reviewer C

This is a retrospective analysis of disparities related to race and a few social determinants of health with treatment access and outcomes among patient with NSCLC. The data is from a single-institution's prospective database and included patients diagnosed between 2000 and 2015. The study is very well-written and is comprehensive in that it evaluates access to care and overall survival. The study reports some findings using variables which are somewhat novel and not widely reported in the literature. Concerns would be that the data are old and that the period reported has been extensively evaluated and reported in the literature. It also has some major limitations related to lacking variables important for evaluating social determinants of health, lung cancer, and survival. Updating the methods and dataset to include patients diagnosed up to 2018 or more preferably up to 2022 would overcome these limitations and would better inform the readers on whether these well-described disparities are persistent.

Authors: We thank Reviewer C and are pleased that this reviewer found the manuscript "very well-written" and "comprehensive," and the figures/tables "Comprehensive and good quality." Please see the detailed responses below to the additional concerns.

METHODS:
Comment #20-Though there is nothing necessarily wrong with excluding patients diagnosed at sites other than IUH, exclusion of patients diagnosed outside of a system then referred into a system for treatment may alter inference of a study that is otherwise designed to evaluate access to health services. I would suggest performing a sensitivity analysis to investigate what impact on the analysis that the inclusion of these patients might have on your results and inference.
Reply #20: We agree that the purpose of this study is to evaluate the impact of specified SDOH on NSCLC diagnosis, treatment, and outcomes, which may inadvertently impact access to health services. We elected to exclude those diagnosed at another site as distance to IU Simon Comprehensive Cancer Center at time of diagnosis may introduce an additional barrier to lung cancer care.
Changes in the text: none.

Comment #21: What is the reason for using old data? The most recent patients were diagnosed >8 years ago. For single institution studies I would suggest using data that is at least more recent than what the large national databases would be able to provide. If at all possible, an updated analysis including patients diagnosed up to 2018 would still allow for ~5-year survival outcomes.
Reply #21: We acknowledge the concern about the dates included in this retrospective study. We agree that including those diagnosed up to 2018 or 2022 could address the concern about the data set having some overlap with previously published data. However, our data includes analyses to try to get to the "root cause" of disparities in lung cancer outcomes. Included in this are novel observations that SDOH, including insurance and marital status, and in some cases, race itself, may impact care decisions in early-stage NSCLC. Racial disparities persist in medicine today. Our concern is that our observations on disparities in treatment are likely to remain unaddressed if providers are not aware of their existence. Further, extending the diagnosis date to 2022, would include COVID-19 as a confounding variable. For instance, Araujo S.E.A. et al found a significant decrease in patients undergoing cancer treatment after the COVID-19 pandemic. Reyes et al. found that 38% fewer new lung cancer cases were diagnosed during the COVID-19 pandemic compared with pre-COVID-19, with more symptomatic and severe NSCLC diagnosed during the pandemic (see below).

(Reyes R., López-Castro R., Auclin E., García T., Chourio M.J., Rodriguez A., López L., Laguna J.C., Lucena C., Molins L., et al. MA03.08 Impact of COVID-19 Pandemic in the Diagnosis and Prognosis of Lung Cancer. *J. Thorac. Oncol.* 2021;**16**:S141. doi: 10.1016/j.jtho.2021.01.219; Araujo S.E.A., Leal A., Centrone A.F.Y., Teich V.D., Malheiro D.T., Cypriano A.S., Cendoroglo Neto M., Klajner S. Impact of COVID-19 pandemic on care of oncological patients: Experience of a cancer center in a Latin American pandemic epicenter. *Einstein.* 2020;19:eAO6282. doi: 10.31744/einstein_journal/2021AO6282.)

Changes in the text: We have added a statement in the methods section, lines 140-145. Please also see Reply #1 for edits including 5-year overall survival.

Comment #22: How were patient comorbidities accounted for?
Reply #22: Patient comorbidities were not included in the database we used and therefore, we acknowledge this as a limitation in the discussion section.
Changes in the text: This was addressed in the Discussion section – 4.4 Strengths and limitations, Lines 532-536 and 542-546.

Comment #23: Though they are both Public, Medicaid and Medicare are very different programs and encompass different age ranges and healthcare access and probably should not be categorized together. For single institutional studies were the majority of patients receive Medicaid from a single state, I would definitely suggest separating Medicaid from Medicare because Medicaid outcomes here could actually be a very specific and novel in that few studies would have this level of granularity. If the Medicaid population is

too small to evaluate alone, it may be more appropriate to combine with the uninsured population as other studies have done (in some states patients diagnosed with cancer are auto-enrolled in Medicaid though I do not know if your state does this).

Reply #23: We agree with the reviewers' comments and have reclassified the insurance groups after reviewing the literature further. Please see the *Reply #6* for a more detailed discussion.

Changes in text: Reasons for not categorizing insurance as "Medicaid/uninsured" vs "other forms or vs "private" insurance are described lines 158-161. Insurance status was recategorized in Table 1 to include the following categories: Medicaid only, Medicare only, Medicare and Medicaid, Self-pay/uninsured, and Private. Subsequent analysis was redone for all analyses, and edits made to the associated results sections and tables involving insurance as a variable.

Comment #24: It appears that some of the models are multivariable (one dependent/outcome variable in model) rather than multivariate (multiple outcome variables in model). These stated outcome variables are likely correlated so multivariate may be appropriate if the sample size is appropriately large. Please confirm which methods were used and change the language if appropriate.

Reply #24: All of the analyses are multivariable and not multivariate.

Changes in the text: This was changed in the text and in the respective tables in the results section.

Comment #25: Consider Poisson or negative binomial regression with IRR rather than logistic regression. The rates of the outcomes appear to be outside of the range for appropriate use of logistic regression models (the outcomes are not "rare" by convention).

Reply#25: This is a very good comment. However, the Poisson regression and negative binomial regression are for count data. None of our outcomes are count data. So, the Poisson regression or negative binomial regression might not be appropriate for our outcome. We have decided to keep the regular logistic regression for our binary outcome and use multinomial logistic regression for the nominal outcomes.

Changes in the text: None.

Comment #26: Suggestion for additional statistical analysis: correlate SDOH with outcomes and investigate collinearity of SDOH variables and race/ethnicity and add supplemental table if deemed appropriate.

Reply #26: We conducted Chi square test of correlation between categorical SDOH and found gender, cigarette use at the time of diagnosis, and marital status were correlated with one another. Gender and marital status being correlated was not surprising. Insurance status, marital status, and race were also correlated. This supports our findings that race alone is likely not the sole driver of some of the observed disparities, but rather those uninsured or underinsured (Medicaid) are more likely to be associated with Black race, compared to White race. Despite the collinearity of these SDOH, we elected to use each of them in our univariable and multivariable analysis. We conducted univariable, multivariable, and KM curve analysis to correlate SDOH with the outcomes of diagnosis stage, 5-year overall survival, stage specific treatment, and reasons for disparities in curative intent surgery in early-stage disease. We included ethnicity in the descriptive table 1; however, we used race as an inclusive term throughout subsequent analysis. For both Black and White race, Hispanic ethnicity made up only 0.25% and 0.58%, respectively, and therefore, subsequent analysis was made using the broader categories or race, since the largest percentage of these two racial groups were non-Hispanic. We appreciate the suggestion of additional analysis; however, we did not make this change as we felt there were other analyses performed in this revision that made the findings of this study stronger.

Changes in the text: None

Comment #27: Are any of the other key SDOH available in this prospective database for analysis? The variables available are unfortunately only a small (though important) aspect of SDOH. Income and education were not available but are there any geographic/neighborhood/Zip Code variables that could be included (if possible) to make this a more robust assessment of SDOH and NSCLC? Suggest reviewing https://health.gov/healthypeople/priority-areas/social-determinants-health for ideas. I understand it may not be possible to expand these given database limitations, but the SDOH variables included are not novel.

Reply #27: We agree with the reviewer that there are SDOH such as education and neighborhood, not available in our cohort, that could be evaluated and are likely important. In fact, an ongoing prospective study including much of this data is ongoing. Many of these have been correlated with SDOH used in this

study. For example, one's educational level is related to one's income level, and insurance status may be considered a surrogate in some instances as those from a lower income status are more likely to be uninsured or have Medicaid, which is based on income level. While zip code was available, we thought it was most important to focus on individual SDOH associated with the observed disparities in diagnosis, treatment, and all-cause mortality as opposed to those generalized based on geographic location. These limitations was addressed in the discussion section.

Changes in text: Discussion section, lines 526-528 and 542-546.

Comment #28: Able to add rurality and/or travel distance as a SDOH? Previous research suggests that travel distance was associated with decreased odds of receipt of adjuvant therapies for NSCLC in both national and single institution data (Logan et al and Eisenberg et al). This may also be true for your institution that likely covers both rural and urban populations.

Reply #28: We agree with the reviewer that rurality and travel distance to IU could represent additional SDOH to be studied. This is a retrospective study and these suggested variables are not available in our registry. We agree these are variables that have been shown to impact NSCLC diagnosis and outcomes and we intend to look at this in future prospective analysis.

Changes in text: No changes made in the text.

Comment #29: For patients who did not receive surgery or radiation, are you able to see if they actually were referred to and saw a surgeon or radiation oncologist? It is unclear who the "gatekeeper" may be here when surgery or radiation are "not part of the plan." This is an important distinction for quality improvement that I hope your dataset can shed some light on.

Reply #29: Unfortunately, we are unable to determine this aspect of care. We agree this is an important factor to consider and in ongoing studies we intend to take a deeper dive into these findings to determine with more granularity actionable interventions that can mitigate the disparities we observed in curative intent surgery of early-stage NSCLC.

Changes in text: No changes made. Will plan to study this in future studies.

RESULTS:

Comment #30: No major comments on the results. The results are very comprehensive and to me there were well-organized. I would suggest adding median overall survival.

Reply #30: We are pleased that you found the results comprehensive and clear. We have revised our survival analysis to include 5-year overall survival and specifically looking at stage at diagnosis, along with our other SDOH of interest. Median overall survival can be extrapolated from the KM curves in figures 1 and 2.

Changes in the text: Results section and Figures 1 and 2.

DISCUSSION:

Comment #31: Key Findings: existence of disparities are already well-documented in prior literature in the time frame that this study encompasses (2000-2015). More recent data for this analysis would better inform the readers if these disparities were persistent despite the work that has been done.

Reply #31: We acknowledge your concerns. As detailed in Reply #21, we do not believe that our findings are minimized by the dates includes, given persistence of disparities to this day. Further, including later diagnosis raises concerns of the COVID-19 pandemic as a confounder on many of our results. We have added a statement in the discussion regarding our rationale.

Changes in the text: We have added a statement in the methods section, lines 140-145. Please also see Reply #1.

Comment #32: Why do you think that your results show Black patients do not have worse overall survival in Cox models despite the reduced access to care?

Reply #32: Supplemental Table 9 suggests that race, a social construct, does not solely account for observed NSCLC overall survival time. Black race is more often associated with worse SDOH, such as being uninsured, which was shown in our study, and the SDOH that disproportionately affect Black individuals are likely the modifiable factors affecting NSCLC disparities.

Changes in the text: Additional survival analysis were done and are discussed in Results, shown in Figure 1 and Supplemental Figure 2.

Comment #33: "It also suggests that implicit bias may impact patient and physician practices pertaining to interventions such as receipt of potentially curative surgery for early-stage NSCLC." This statement makes me wonder if these patients are actually being referred, because in practice virtually all patients with early stage disease will be offered curative surgery if seen by a surgeon and are deemed operable. How could future research determine where the disconnect in this process is?"

Response #33: This is a good question. Future studies could assess referral rates of those diagnosed with stage I NSCLC, if they are seen by a surgeon, and then if surgery was offered. This is one of our ongoing studies now where patients newly diagnosed with Stage I NSCLC are being asked if they were referred to a surgeon, and if not, what was the reason this was not recommended as an option.

Changes in the text: Discussion, lines 471-474.

Reviewer D

Comments to the authors Duncan et al on "Racial Disparities in Staging, Treatment, and Mortality in Non-Small Cell Lung Cancer (NSCLC)."

Duncan al present data from a single center patient database on patients with NSCLC diagnosed in a region in the USA. The study includes >3,300 patients with lung cancer diagnosed in 2000-2015 and the study aims to examine differences in stage at diagnosis, treatment and survival in patients with Black race compared to White race investigating also the influence of social determinants of health. The research question is of interest as equality of health care is a frequently discussed topic and identifying potentially targetable focus areas may improve health care equality over time. The authors show variations in treatment and survival according to race and can with regression analyses prove that these differences can be explained by social determinants such as insurance type which may affect different aspects of future prevention-programs and directed health care support-programs.

There are several strengths of the study: selection bias is kept to a minimum, the relative comprehensiveness of data and the manuscript is very well written and easy to follow. Figures/tables are adequate. The major limitations are discussed.

I still have some comments for the authors and editor to consider:

Major comments:

Comment #33:. Staging was categorized into two entities, early stage (I-II) and advanced stage (III-IV). I advise against including Stage IIIA to the group with stage IIIB-IV. Whereas Stage IIIB-IV commonly are grouped together and largely receive similar treatments, stage IIIA is very heterogeneous group and a significant proportion receive potentially curative treatment. Long term overall survival far exceeds that of stage IV in this category even among those that progress on curatively intended therapy.

Reply #33: We considered the reviewer's comments regarding differing treatment options for stage IIIA and stage IIIB-IV, and, in some cases, different outcomes. We felt the broader categories, early (stage I and stage II) vs advanced (stage III and stage IV) classifications were still appropriate as this is the reason we evaluated multivariable analysis for only surgical resection in early-stage disease and only chemotherapy in advanced-stage disease (as each person in these classifications should have received these treatments, respectively). In survival analyses, we re-analyzed the data based on separated stages to account for the shorter survival time of those diagnosed with later stage disease.

Changes in the text: Refer to result section and Figure 1.

Comment #34. Treatment was divided into four categories: surgery alone, radiation alone, systemic therapy and systemic therapy plus surgery. Here, the foremost used treatment for Stage IIIA, curatively intended chemoradiotherapy is missing. This is important both from the clinical perspective when analyzing outcomes but it is also is an avenue of potential inequalities in care. Chemoradiotherapy is resource intensive, costly and is often limited to specialist centers, especially in the earlier time period of the study cohort. Access to this potentially curative treatment is an important aspect to analyze given the subject of the paper.

Reply #34: We agree with the reviewer's comments regarding curative intent chemoradiation. We have recategorized treatment options in table 1 to surgery only, radiation only, and systemic therapy.

Changes in the text: Results section. Table 1.

Comment #35: For an international audience it would be prudent to give some background on the American health care system. This could be provided in the introduction section or with a suitable reference. For example in section 1.2 Medicaid is not explained to be the public insurance – that is explained only later on in the text in section 2.2.1. Also, the location of the study in the methods section is only described by city, country is also not stated in the author's affiliations on the title page. I suggest that you include that the study was performed in the USA.

Reply #35: We appreciate the reviewer's comments regarding clarification for an international audience. This has been added to the affiliation and methods sections. An explanation of Medicaid and Medicare as a differing public insurance types level has now been included in the text in the Methods section, and data has been analyzed based on further stratification of the specific type of public insurance.

Changes in text: The location of the United States of America has been added to the affiliations on the Cover Page, as well as Methods section, line 133. A description of U.S. public insurance types has been added to lines 97-103.

Comment #36: In northern Europe the median age at diagnosis of NSCLC is 69 years. Here only 28% have an age of 71 or older. Is this representative of the population in the hospitals uptake area or do you have any other explanation for the relatively low age at diagnosis?

Reply #36: In the US, the median age of lung cancer diagnosis is 70 years of age. While we are not certain to the exact cause of the lower percentage of this cohort being diagnosed with NSCLC age at age < 71, we do believe that cigarette smoke could be contributing to this difference. Indiana ranks 10th in the U.S. for the number of adults who smoke, and 13th in the highest number of smoking related deaths in the U.S; therefore, cigarette use may contribute to diagnosis at a younger age. Black individuals also tend to be diagnosed at a younger age, and this could contribute to the overall lower diagnosis age of the group. https://wwww.in.gov/health/files/IndianaTobacco_Burden_7_12_11.pdf. https://www.lung.org

Changes in text: none. Age is defined in Methods, lines 168-169.

Comment #37: Minor comments:
1. In section 3.3., paragraph 2, row 224 space is lacking between "with private insurance (OR….)"
2. In section 3.4 row 243 space is lacking between "95% CI"
3. In section 3.8, paragraph 2, row 291 the square bracket before 95% is with another symbol.

Reply #37: Thank you for finding these grammatical errors. The requested changes have been made.

Changes in text: changes made, sections 3.3, 3.4, and 3.8, as recommended

Reviewer E

I commend the authors on investigating a known and unfortunate truth of cancer outcomes, racial disparity. The authors have sought to explore the role of racial disparities through social determinants of health in the outcomes of patients with non-small cell lung cancer. They are some of the first investigatory to also include patients presenting with late disease and not just early-stage cancer in their analysis. The manuscript is well written. Please see below for my suggestions by section.

INTRODUCTION:

Comment #38: Within the 1.2 Rationale and knowledge gap section please briefly mention the standard of care treatment options for early-stage NSCLC with references (ie curative surgical resection vs SRT etc). Although it is important to note that there are disparities seen in who undergoes upfront curative surgical resection there are many factors that could contribute to this, including the robustness of the cardiothoracic surgery program at the institution. You mention this briefly in the covariates section but references would be nice.

Reply #38: We have added text within the introduction section along with references discussing the standard of care treatment options for early-stage NSCLC. We also provided a reference for how surgeon practices can influence treatment decisions for early-stage NSCLC.

Changes in the text: Introduction section, lines 108-110 (with additional references) and methods, 178-179. Discussion section, lines 471-474, includes a discussion of the potential impact of differing surgical practices.

Comment #39: METHODS:

I am somewhat concerned by the timeline of inclusion (2000 to 2015) and the inclusion of all AJCC editions over the period. It would require too much work to retrospective update patients AJCC staging to edition 7 when considering approximately 3400 patients. Also, this would have more implications for late stage disease. This would be important to mention in section 4.4 strength and limitations.

Reply #39: We agree with the reviewer that studies analyzing the effect of transitioning from the 6th to the 7th edition AJCC staging have shown that using the later staging edition could lead to upstaging, and presumably affect survival analysis. However, one such study only included those who had undergone surgery, and therefore did not include those who were non-surgical candidates and those with advanced stage disease (ie stage 3B and 4). We have included this as a limitation of our study. We elected to not pursue this as we wanted to assess how the SDOH impacted staging and treatment decisions at that time based on the current staging guidelines used. (Erdoğu V, Çıtak N, Sezen CB, Aksoy Y, Aker C, Doğru MV, Emetli EY, Onay S, Saydam Ö, Metin M. Comparison of 6th, 7th, and 8th editions of the TNM staging in non-small cell lung cancer patients: Validation of the 8th edition of TNM staging. Turk Gogus Kalp Damar Cerrahisi Derg. 2022 Jul 29;30(3):395-403. doi: 10.5606/tgkdc.dergisi.2022.20089. PMID: 36303702; PMCID: PMC9580290.)

Changes in the text: Discussion Section 4.4, lines 550-556.

Comment #40:

How was lost to follow up handled among your study? I am assuming there were patients that either were diagnosed at your institution and then received care elsewhere, even in another state/city (private insurance) and also people that were lost even after receiving treatment.

-Similarly, given overall survival is an outcome how was this assessed? For example was a social security database utilized? However were patients lost to follow up and without proof of death assess in the study?

Reply #40: Patients diagnosed at our institution who received care elsewhere were excluded from this study. While these may be referred from other institutions, the cohort only included those newly diagnosed with NSCLC at IU Simon Comprehensive Care Center who subsequently received their care here as well. Regarding survival analysis, a SSN database was not utilized. Instead, date of death or date of last contact was used when for overall survival analysis. This is included in the Methods section of our manuscript.

Changes in the text: This is explained in Methods, lines 163-165 and line 211.

Comment #41:

RESULTS: It is mentioned in the results, specifically section 3.4 regarding patients who failed to receive curative intent surgery, that one reason was "contraindications due to other conditions". I do not see in basic demographics a description of common comorbidities that might preclude patients from having surgery such as severe COPD, CHF, etc. This may be helpful to also see if there is a difference in the baseline comorbidities between race/ethnicity as has been reported before.

Reply #41: We agree with this reviewers' comments that "contraindications due to other conditions" was not specified in the manuscript, nor was this detailed in our dataset. We have included mention of this as a limitation in the discussion section, and have ongoing investigations based on our findings to determine specific causes that guideline-appropriate treatments were not received in early-stage NSCLC.

Changes in the text: Discussion section, lines 532-536, 542-546.

Comment #42: Regarding section 3.5, it states receipt of radiation was significantly associated with cigarette use. Does your center or the CT surgery department not perform surgery in current tobacco users? This can be the policy for certain providers etc but of not would be something to mention in the methods.

Reply #42: While tobacco cessation is highly recommended prior to surgery, we are not aware of any policy/ practice that prohibits surgery in those who smoke cigarettes.

Changes in text: none

Comment #43:

Regarding multivariate analyses. I noticed that the analyses had variables which contained black vs white in them, such as race insurance, race stage at diagnosis… if these were included with variable "race" usually listed at the top of the table it could be a confounder and why some of the race variables themselves were not statistically significant.

Reply #43: We removed all interaction terms in our models, as they were not statistically significant.

Changes in text: Results section and Supplemental Table 8 in the manuscript.

Comment #44:
Was tumor grade assessed on univariate analysis to be associated with overall survival? Given this known associated with survival in lung cancer literature this should perhaps be included in table 9 multivariate analysis for overall survival.
Reply #44: We did not assess tumor grade and survival. We have listed the number and percentage of various tumor grades in Table 1 demographics; however, the larger percentage, 55.06%, did not have their tumor grade information, so we did not perform subsequent analysis on this variable.
Changes in text: None.

Comment #45: CONCLUSION:
Try and abbreviate and make section more concise.
Reply #45: Thank you. Edits from the conclusions are included as a separate manuscript
Change in text: Multiple. See extensive edits to the discussion.

Reviewer F

This is a well-written and detailed study describing the impact of SDOH on stage at diagnosis, receipt of stage appropriate treatment or reasons for non-receipt, and mortality. They specifically evaluated if the impact of race on these outcomes is mediated by various SDOH. Key findings indicate that Black compared to White race was associated with advanced stage, less curative-intent surgery, receipt of systemic therapy in both univariate and multivariate analyses. There were no race differences in receipt of radiation or time to treatment, and Black race had worse overall survival in univariate analyses but not when controlling for other SDOH. A key strength of this study was the ability to simultaneously account for many SDOH and consider their independent association with outcomes and the extent to which they collectively mediate the impact of race on outcomes. Comments for consideration are as follows:

Comment #46: The methods indicate that 'unknown' indicates and includes missing registry data for SDOH and covariates. If this is the case, why are there difference 'N's for some of the variables in Table 1 (eg gender, ethnicity, tumor grade, marital status)? Please clarify.
Reply #46: We understand this was confusing in the initiation manuscript so we have clarified this in this current version.
Changes in text: Lines 163-165. We have added "Ns" to Tables.

Comment #47: Was there any exclusion of patients who died within a short timeframe post-diagnosis (ie didn't survive long enough to initiate treatment), or are these patients as well as those who were diagnosed at autopsy included in the analysis? If so, consider excluding these so this won't bias results regarding mortality and receipt of treatment, or at least report the proportion meeting this criteria.
Reply #47: The number of patients that died within the first 2 months of diagnosis was 212. Survival analysis was similar when we included and excluded the 212 patients that died within the first two months. Since one of our outcomes of interest was the impact of SDOH on staging and overall survival, we felt it was important to include those who died within the first 2 months as these SDOH could be potential identifiers of those with worse overall survival.
Changes in text: New Kaplan-Meier survival curves (Figures 1 and Supplemental Figure 2) and Results section 3.8, lines 364-367.

Comment #48: For clarity, provide 'N' in Tables 3-6 (not just in the supplemental tables) since these are stage-specific subsets of the population?
Reply #48: We have updated tables 3-6 with "N" as requested.
Changes in text: Results section, Tables 3-6.

Comment #49: What's the 'N' for Supplemental Table 7 since it's obviously limited to those who had some type of treatment? The 'treatment type' variable in Table 1 doesn't give an indication of how many patients did not receive any treatment.

Reply #49: We have included the "n" for supplemental table 7.
Changes in text: Supplemental Table 7. Tables 3-6 as well.

Comment #50. While evaluating the combined effect of SDOH is informative, it would be helpful to assess the potential effect modification of each SDOH (eg via interaction terms) to determine if there are key SDOH mediators on the association between race and the outcomes of interest.
Reply #50: We checked for interaction terms among the SDOH and there was no statistical difference. Multivariable analyses were included for each intervention, which we believe better recapitulates the interactions between race, SDOH and outcomes. However, we acknowledge different ways to measure these interactions.
Changes in text: None