Peer Review File

Article information: https://dx.doi.org/10.21037/ace-22-1

Reviewer Comments

Comment 1: It was not clear what level of details are available in the dataset about the cause of mortality.

Reply 1: We used the "SEER cause-specific death classification" variable from the NAACCR Cancer in North America (CiNA) Deluxe data file in the cancer-specific mortality analyses. This variable categorizes cases as:

- "Dead (attributable to this cancer diagnosis)"
- Alive or dead of other cause
- Dead (missing/unknown cause of death)

This variable takes into account cause of death information (ICD-10 codes), site of original cancer diagnosis, tumor sequence, and diseases related to the cancer of diagnosis.

Changes in text: We have clarified these details in the Methods and added a reference for this variable (see Page 7, lines 122-128).

Comment 2: Did you adjust for treatment modality for the models estimating late-stage diagnosis? If so, why?

Reply 2: We clarify that we only adjusted for treatment modality in the mortality models.

Changes in text: We have clarified these details in the "Statistical Analysis" subsection of the Methods (see Page 8, lines 150-152).

Comment 3: Since you have several outcomes and different sample sizes, it will be extremely helpful for the readers to see a table to help clarify all the models you ran – sample, outcome, stratification, interaction, adjusting covariates, type of model.

Reply 3: Thank you for the feedback. We have created a table outlining the recommendations above.

Changes in text: This table is now included as Supplemental Table 1 and referenced in the Statistical Analysis section, p. 8, line 180.

Comment 4: Please provide reference for the CDC methodology you used to identify HPV-associated cancer cases from the dataset. Has this been validated in previous studies? The link in bibliography is inactive I think the new link is https://www.cdc.gov/cancer/uscs/public-use/pdf/predefined-seer-stat-variables-508.pdf. Also, studying the methodology on this link, what you calculate are attributable fraction. When you multiply it by associated cancers, you get attributable cancers. I do not see in the paper how you ascertained HPV attributable cancers.

Reply 4: Thank you for the feedback. Our aim was not to estimate attributable cancers due to HPV infection. We identified our HPV-associated cancers using the

most recently updated (06/8/2021) version of the CDC Definitions of Risk Factor-Associated Cancers, and included the correct link: (https://www.cdc.gov/cancer/uscs/public-use/predefined-seer-stat-variables.htm). This is described in the "Data Source and Study Sample" subsection of Methods starting on p. 5. These same methods have been previously used to conduct cancer registry-based studies on HPV-associated cancers. Additionally, we have acknowledged the lack of molecular confirmation of HPV status as a limitation in our "Discussion" section (p. 14, lines 309-314 when considering that cancer registries unfortunately do not collect HPV status as part of data extraction for cancer cases.
Changes in text: We have cited studies that have used similar methodologies in the "Data Source and Study Sample" subsection of the Methods (see Page 6, lines 108-109).

Comment 5: You don't explain in methods, the categories of treatment modality. You adjusted for both, the initial and reevaluated treatment modality? Weren't they correlated? Please explain the categories of treatment modalities you mention in your tables.
Reply 5: We clarify that the treatment modality was only for the first course of planned treatment. For oropharyngeal cancers, there are generally three treatment modalities, which include surgical removal, systemic chemotherapy, and radiation therapy. Either of these treatment modalities can be used individually or in different combinations with one another. We categorized the treatment modalities based on what was reported to the cancer registries, whether they had no treatment, one type of treatment, a combination of modalities, or all available modalities. The categories included: surgery only; radiation or chemotherapy only; surgery plus (radiation or chemotherapy); radiation and chemotherapy; all modalities; or no treatment.
Changes in text: We have clarified these details in the "Covariates" subsection of the Methods (see Page 8, lines 150-152).

Comment 6: Did you merge the dataset with another dataset at county level, to identify county-level % of people living below FPL?
Reply 6: We did not merge the dataset as the county level data is provided by NAACCR in the CiNA deluxe dataset. Information on how specific county attributes were calculated can be found at: https://seer.cancer.gov/seerstat/variables/countyattribs/static.html#15-19.
Changes in text: We clarify that all variables analyzed were available in NACCR CiNA (see Page 6, lines 96-97).

Comment 7: On page 7, please explain what are multivariable model 1 and 2 – why is stage at diagnosis on the right side of equation, from earlier description it seemed, it was the outcome. Again, I think a lot of these questions will be answered if you include a table to explain your models.

Reply 7: Thank you for this feedback. Mortality multivariable model 1 included age at diagnosis, health insurance, county level attributes of residence (Metropolitan/Non-metropolitan, percent of persons below poverty); geographic region of the U.S., and stage at diagnosis. We included treatment modality in multivariable model 2 to determine if treatment would have any effect on the hazard ratios observed in model 1.
Changes in text: We have included Supplemental Table 1 based on Comment 3 and clarify the covariates used in the cancer-specific mortality models (see page 7 line 137) as well as Models 1 and 2 (Page 8-9, lines 163-175).

Comment 8: If you have multiple outcomes, you should adjust your p-value significance threshold to account for chance using Bonferroni or other adjustment methods.
Reply 8: The outcomes used in our models in this study are independent of each other. We are not performing multiple comparisons; therefore, we believe that adjusting using Bonferroni correction is not necessary.
Changes in text: None.

Comment 9: On Page 12, you say one of the limitations is that you might have included non-HPV cancer cases because HPV status was not molecularly determined. However, On Page 10, paragraph 2, towards the end you explain that tobacco consumption is higher among NH black males than white males, and that is a risk factor for OPC. So, my concern is that without molecular confirmation of HPV status, it is likely that among Blacks and possibly Hispanics, you might be analyzing OPC cases that are not HPV positive, and hence clinically different from HPV positive OPC cases, which as we know are better manageable and have better survival. So, some of the bad survival and mortality outcomes that you observe in your analysis among NH Blacks and Hispanics could be attributable to their non-HPV status? How do you reconcile this and feel confident that you are comparing apples to apples across different racial-ethnic groups? And this ties to one of my earlier comments about calculating attributable cancers rather than attributable fraction.
Reply 9: Thank you for this feedback. We understand the concern of the reviewer which is why we acknowledged these limitations in the Discussion starting p. 14, line 301. Admittedly, these are some of the inherent setbacks in using cancer-registry data in conducting this type of study. We acknowledge that compared to NH Whites, NH Blacks have higher smoking rates and Hispanic males have lower smoking rates, which could be confounding the relationship between race/ethnicity and OPC survival. Despite this limitation, the NAACCR datafile is the most comprehensive database for conducting epidemiologic studies on oropharyngeal cancers and does provide us with useful insights, particularly for HPV-related studies among males and race/ethnic comparisons. Unfortunately, neither molecular HPV status nor smoking behavior was available in the NAACCR CiNA dataset as they are not collected by cancer registries for OPC. Further

research is needed to confirm our findings using molecularly confirmed HPV status and controlling for other risk factors for OPC such as smoking.
Changes in text: We have clarified these details in the Discussion (see page 14, lines 305-314).