Peer Review File

Article information: https://dx.doi.org/10.21037/jtd-23-428

Reviewer A

Using 2004-2018 SEER data, Liu et.al examined the survival differences between surgery and radiotherapy among elderly (≥70 years) small cell lung cancer (SCLC) patients. The authors found that SCLC patients who received surgery had significantly better survival than those receiving radiotherapy. The manuscript has following flaws in study design, data analysis, and writing.

1. The guidelines recommend radiotherapy combined with adjuvant or concurrent chemotherapy as an alternative treatment to surgery. However, this study compared the effectiveness of surgery vs. radiation therapy alone, but not surgery vs. radiation therapy + chemotherapy. It is not surprising that surgery is associated with better outcomes than radiation therapy alone, but this finding is less clinically relevant without considering concurrent or adjuvant chemotherapy with radiation therapy.

Reply 1: Thank you for your comment. We agree that comparison of concurrent or adjuvant chemotherapy with radiotherapy will be more clinically relevant. However, information on chemotherapy was unavailable in the SEER database. That will be a major part in our future research using data collected from our institution.
Changes in the text: no changes were made in our text.

2. Comorbidity is a critically important confounder in this association, as patients who have less comorbidity tend to receive surgery and better survival. So, the observed association is very likely due to the confounding effect from comorbidity. Without adjusting for comorbidity, this conclusion is not valid.
Reply 2: Thank you for your comment. We agree that adjusting for comorbidity will be more clinically relevant. However, information on comorbidity was unavailable in the SEER database. As age is considered as a determinant of comorbidity, thus we regard age as a proxy of comorbidity to offset this limitation.
Changes in the text: no changes were made in our text.

3. It is not clear how IPTW, the major statistical analysis of this study, was applied. Using propensity score?

Reply 3: Thank you for your comment. Inverse probability of treatment weighting (IPTW) uses the estimated propensity score to balance baseline covariates between different treatment groups.
Changes in the text: no changes were made in our text.

4. Given the study period is very long (more than a decade), secular trend should be considered, i.e., year of diagnosis can be adjusted in the model.

Reply 4: Thank you for your comment. Secular trend of local treatment modality has been evaluated in our study. The model including year of diagnosis will be a part of our future study.
Changes in the text: no changes were made in our text.

5. Insurance is available in the data from 2007. The authors can conduct a sensitivity analysis adjusting for insurance with data after 2007.

Reply 5: Thank you for your comment. Since we enrolled patients from 2004 to 2018 in SEER database, insurance data was unavailable for a small percentage of patients. The sensitivity analysis adjusting for insurance will be a part of our future study.
Changes in the text: no changes were made in our text.

6. There are many grammatical errors in the manuscript. Writing needs to be improved.

Reply 6: Thank you for your comment. We feel sorry for the grammatical errors in the manuscript, we had improved the language using a language editing service.
Changes in the text: no changes were made in our text.

Reviewer B

In this manuscript, the authors analyzed that the difference of survival in elderly patients with early-stage SCLC using IPTW propensity-score analysis in a data-based cohort. Surgery was associated with a longer OS time than radiotherapy in all patients in their analysis but no difference was observed in the 5-year cumulative incidence rate of cancer-related death between the surgery and radiotherapy groups in patients aged ≥80 years.
The following concerns should be cralyfied before publication.

#1
Most crucial question is 'Which is better treatment: surgery with chemotherapy or chemoradiotherapy?' because chemoradiotherapy is one of the established, recommended treatments in young patients with early-stage SCLC.
The analysis of overall survival of four-arm comparison of surgery alone, surgery+chemotherapy, radiation alone, and chemoradiotherapy is required to clarify the best treatment option in this population.

Reply 1: Thank you for your comment. We agree that comparison of four-arm including surgery alone, surgery+ chemotherapy, radiation alone, and chemoradiotherapy will be more clinically relevant. However, information on chemotherapy timing was

unavailable in the SEER database. That will be a major part in our future research using data collected from our institution.
Changes in the text: no changes were made in our text.

#2
In page 7, line 6 and 20, page 8, line 1, the authors describe as ‘modern staging and treatments,’ and ‘modern radiotherapy techniques.
The term of modern is vague in this discussion. The authors should clarify the definition of what they mean by ‘modern’ staging and treatment, or ‘modern’ radiotherapy because the authors enrolled the patients with SCLC in relatively old start year and in long time period (from 2004 to 2018).

Reply 2: Thank you for your comment. The term modern staging and treatment mainly indicated the application of AJCC staging system. The definition of modern treatment cannot be made as information regarding chemotherapy regimen or radiation technology were unavailable in SEER database.
Changes in the text: no changes were made in our text.

Reviewer C

This is an interesting study, but is inconclusive, I think the authors overstate what can be concluded.
Overall, I think the design and writing is good. My problem is a misunderstanding of what we can get from propensity matching. It is not magic, and the way we apply it often leads us away from being true to science. We have to be thoughtful.
A prerequisite for PM is that all known or suspected factors are included, or at least those arguments can be made that ones that are missing are very unlikely to be confounders. Unfortunately, this is almost universally ignored in clinical medicine studies. The authors are making the same mistake that essentially everyone else makes as well. Performing a Cox analysis to identify prognostic factors doesn’t get you off the hook if you don’t include potential factors in the cox analysis.

The way I approach conducting such studies, or assessing such publications, is to start with a theoretical list of potential confounders. Top of the list in this case would be comorbidities (including severity of comorbidities) and functional status (e.g., PS). I would also wonder about socioeconomic factors, and access (region, rural/urban, insurance). Then I would wonder about anatomic location (not which lobe, but central peripheral). Size (or T stage) would also be there. Of this list of most likely factors, the study has only accounted for 1. A further question that is raised is the time period, given the shifting rates of surgery and RT. I think this should also have been included, although I don’t have data that this is a likely factor.

Given the known predisposition (in fact even GL recommendation) that more robust patients be treated surgically and less robust patients with RT, I think it is highly possible that the outcome differences seen are due to this confounder and not the treatment received. The role of chemo seems to underscore this – less robust patients are less likely to get chemo in addition to a local therapy.

The way I assess PM studies is not by looking at a table of how well match a select number of factors are, it is by assessing what factors are not included and how important I think they are.
I also think that we, in the medical literature, are losing sight of the scientific principle, and thus often mislead ourselves. True science demands that we are skeptical, and try critically to pokes holes into our interpretation, and only when we are unable to do so, accept the interpretation. In medicine, all too often we start with a hypothesis, find data to support it, and then conclude that that we have proven the hypothesis without ever considering possible alternative explanations and then addressing them. The limitations paragraph has become essentially useless in medical literature. It is generally approached as a quick listing of a few limitations, without any exploration of whether they are likely or not to have affected the results and in which direction. That is what a discussion should be – and I think the thing that leads us to true insights.

I don't mean to be harsh, and I acknowledge that this paper, as written, is on par with 99% of such adjusted database studies that are published. I just think we should strive for a higher standard.

In conclusion, I think the interpretation needs to be toned down. I think that a PM study should explicitly list the factors that were included (I assume it is the ones in the table, but perhaps there are more). I also think the discussion needs to critically consider alternative explanations for the observations.

Reply 1: Thank you for your comment. The potential confounders as suggested will be collected in our future research using data from our institution.
Changes in the text: no changes were made in our text.