TRANSLATIONAL LUNG CANCER RESEARCH

Peer Review File

Article Information: Available at http://dx.doi.org/10.21037/tlcr-20-469

Reviewer A

Review "Stereotactic body radiation therapy for empirically treated hypermetabolic lung lesions: a single-institutional experience identifying the Charlson score as a key prognostic factor"

This paper described 90 patients with early stage NSCLC treated with SBRT without a pathology proven diagnosis. The results consisted of local control, distant-free survival and overall survival. The authors state that OS is mainly dependent on the overall health status of a patient reflected by the Charlson co-morbidity index.

Reply 1: This is an accurate description of the paper.

Changes to the text: None.

The paper is well-written and the topic is of interest, although not new. I am not convinced that this paper gives new insight to the already existing literature about this topic. In addition, I have some major concerns about the selected patient group, the treatment itself and the analysis. Therefore I decided to advise to reject the paper.

Reply 2: We believe that this paper uniquely contributes to the literature concerning this topic by identifying the Charlson score as a critical prognostic factor and discussing this result within the context of a relevant BED threshold, lesion size, and pre-treatment SUV. Together, these factors allow for risk stratification of patients under consideration for empiric lung SBRT. Further, the identification of a specific Charlson score threshold in the context is, to our knowledge, the first such threshold that has been identified in the radiation oncology literature concerning such a patient group. We have made major improvements to the manuscript, as detailed by the point-by-point responses, below.

Changes to the text: See below.

TRANSLATIONAL LUNG CANCER RESEARCH

Major comments:

The study population is very small and about 11 patients per year were treated. It would be interesting to know more about the total group of patients treated with SBRT for NSCLC. Is the study group a minority in your patient population or is it quite common? It is important for the reader to know whether the patient group is a reflection of the reality. How many patients in total were treated during 2008-2016 using SBRT?

Reply 3: We agree that this point requires additional description. This study group is, in fact, a minority of the patient population at this institution. Over this time period, there were approximately 750 SBRT treatments to the lung. Of these, about 300 cases involved metastatic disease to the lung from a different primary tumor, and only approximately 4% of the treatments involved small cell lung cancer. Therefore, there were approximately 420 cases of SBRT for NSCLC from 2008-2016, and 91 of these met inclusion criteria for this study.

Changes to the text: 91 such treatments…met all inclusion criteria "from an initial set of approximately 420 cases of SBRT for NSCLC" (see page 10 lines 176-177).

The patient group has a median age of 77 years; the use of an age-adjusted Charlson Co-morbidity Index might change the results.

Reply 4: While we agree that this is a possible avenue for further study, we did not elect to use the age-adjusted Charlson comorbidity index because age itself did not reach statistical significance for either local control or overall survival on Cox proportional hazards analysis. Furthermore, the error bars for age's potential influence were very small, indicating that even if age were to play a statistically significant role in predicting either outcome, its role would be small. We have added a sentence to the limitations paragraph indicating that the use of an age-adjusted score may provide a different angle for additional study.

Changes to the text: Finally, the use of the use of an age-adjusted Charlson Comorbidity index may provide a different angle for additional study (see page 20 lines 411-412).

The technique that was used to treat the patients with SBRT consisted of 3D. IMRT or VMAT was considered but it is not clear why this was not chosen. Could you elaborate on this? How many patients were treated with 3D and IMRT/VMAT?

Reply 5: IMRT and VMAT were considered but not used in this context. This point has been clarified in the text. In this setting, the treating providers found improved dosimetric outcomes using a 3D conformal treatment planning approach, hence its use in this context.

Changes to the text: Intensity-modulated radiotherapy and volumetric-modulated arc therapy were considered "but not used in this context" (see page 11 line 208).

A quarter of the study group refused a biopsy. Could you elaborate on this group, e.g. was it possible to biopsy these patients?

Reply 6: This is correct. Patient refusal of a biopsy was recorded as a reason for a biopsy not being obtained in cases when a biopsy was deemed to be feasible and medically indicated, but the patient elected not to undergo the procedure. This point has been added to the manuscript.

Changes to the text: "In the cases of patient refusal, a biopsy was deemed to be feasible and medically indicated, but the patient elected not to undergo the procedure" (see page 14 lines 255-257).

Is it possible to update the follow-up since the patients were treated up to 2016? Or to expand the number of patients treated until 2019?

Reply 7: Unfortunately, this was not possible; however, the statistical analysis was further developed. Kaplan-Meier estimates for 24-month and 36-month local control, overall survival, and progression-free survival were added. Further, 36-month overall survival estimates were provided to help distinguish between the two key stratification developed in this study (BED $\geq$ 120 Gy and Charlson score $\geq$ 9). Additionally, the figures were improved such that they demonstrated local control, progression-free survival, and overall survival for 60 months, as this is likely a more accurate description of the data. Finally, though the relatively short follow-up was already included in the limitations section of the discussion, that aspect of the paragraph was further developed.

Changes to the text:
36-month local control of 91.3% was achieved. 24-month overall survival and

progression-free survival were 65.4% and 56.0%, respectively...with 36-month overall survival of 50.5% for patients with BED ⩾ 120 Gy and only 31.6% for patients with BED < 120 Gy (see page 4 lines 57-61).
2-year overall survival via the Kaplan-Meier method was 65.4%, and 2-year progression-free survival was 56.0%. 2-year local control was 93.1%, and 3-year local control was 91.3% (see page 15 lines 284-286).
36-month overall survival was 0% for patients with Charlson score ⩾ 9, and it was 50.3% for patients with Charlson score < 9. Similarly, 36-month overall survival was 50.5% for patients with BED ⩾ 120 Gy, but it was only 31.6% for patients with BED < 120 Gy. On the other hand, BED ⩾ 100 Gy failed to demonstrate improved overall survival on Kaplan-Meier analysis (p=0.45) (see page 16 lines 303-311).
Even so, more extensive follow-up might have allowed for further important conclusions to have been drawn (see page 20 lines 408-410).

I was wondering whether an OS benefit was also observed with a threshold BED ⩾100Gy? Onishi stated in 2007 that to achieve local control of 90% a BED ⩾100Gy is needed (PMID 17603311), which was confirmed by several other publications. Why did you choose BED ⩾120Gy?

Reply 8: We agree that this distinction merits discussion. It was added to the results section that BED ⩾ 100 Gy failed to demonstrate improved overall survival on Kaplan-Meier analysis, whereas BED ⩾ 120 Gy trended towards improved survival on Kaplan-Meier analysis and demonstrated statistical significance via other statistical methods.
This distinction was further explored in the discussion section. The study mentioned (PMID: 17603311) was added as an additional reference, and it was stated that this study may have failed to demonstrate a difference at the threshold of BED ⩾ 100 Gy due to the few treatments involving such low doses (only six). It was recommended that further study help distinguish between these two thresholds.

Changes to the text: On the other hand, BED ⩾ 100 Gy failed to demonstrate improved overall survival on Kaplan-Meier analysis (p=0.45) (see page 16 lines 310-311).
Also: Interestingly, this is a higher threshold than previous studies noting improved local control with BED ⩾100 Gy (46). Treatments with BED ⩾ 100 failed to show improved overall survival in this study, but this may be due to the

low utilization of such low dose treatments since only 6 treatments involved BED < 100 Gy. This distinction may be a potential avenue for further study (see page 19 lines 377-381).

To evaluate patients after SBRT using FDG-PET is not commonly used since high dose radiotherapy causes inflammation resulting in elevated SUV-levels. Is it necessary for the analysis to include this?

Reply 9: The authors agree that it is true that high dose radiotherapy can result in inflammation and result in elevated SUV levels. Even so, post-treatment SUV decreased significantly after SBRT (median 1.5 vs. 4.5). Further, we felt that since the hypermetabolic nature of the lesion was a key pre-treatment criterion, tracking this metric may be of value. Post-treatment SUV has even been correlated with local control in some studies (e.g. PMID: 29296342). Additionally, since the follow-up scans are undertaken 2-3 months after completion of treatment, inflammation levels secondary to the radiation itself have generally decreased by the time that follow-up imaging is performed.

Changes to the text: Post-treatment SUV was recorded when available ", despite not being a standard follow-up measure at all institutions" (see page 12 lines 225-226).

The introduction is quite extensive. My suggestion is to shorten it and describe the reasons why you have undertaken this analysis more to-the-point.

Reply 10: We agree that the introduction is too long. We have worked to shorten the introduction while minimizing the loss of important content. Further, we have clarified the purpose of the study in the final paragraph of the introduction.

Changes to the text: Various changes were made to shorten the introduction (see pages 6-9 lines 73-155).
Also: "By considering a range of factors, including biologically effective dose (BED), pre-treatment SUV, the Charlson score, and lesion size, we hope to delineate a set of prognostic factors to assist patients and providers in shared decision-making regarding the utility of empiric SBRT for NSCLC. In particular..." (see page 9 lines 149-152).

114: stage III NOT excluded? Why and tell us more about these 7 patients. My

suggestion is to exclude these patients since SBRT is not used in clinical practice in this patient group.

Reply 11: A thorough description of why stage 3 tumors were included in the study was added to methods section. Some patients with T stage 1 or 2 disease and nodal spread are treated with SBRT when other (more commonly used) treatment approaches are refused. For this reason, they comprise a non-trivial portion of this institution's empiric SBRT. Patients with T stage ⩾ 3, however, were excluded from the study because these tumors are generally too large to be treated with SBRT.
A separate paragraph was added to the results section describing the outcomes and characteristics of patients with stage 3 disease. 4 of the seven patients refused the biopsy, and the median Charlson score was 9. Even so, none of the patients experienced local failure after a median 112.5 Gy was delivered, and only one patient reported further disease progression.

Changes to the text:
Though SBRT is often not the standard treatment for stage 3 NSCLC, at this institution, some stage 3 patients refused chemotherapy or other treatment options and were treated with SBRT. These patients were generally T stage 1 or 2 disease with significant medical comorbidities and nodal spread of disease. Since they comprise a not-insignificant fraction of empiric SBRT, they were included in the study. Patients with T stage ⩾ 3, however, were excluded because these tumors are generally too large for the use of SBRT (see page 10 lines 168-174).
Also: Patients with stage 3 disease were also considered separately, but results were generally similar to the other patients in the study. 4 of the seven patients refused the biopsy, and the median Charlson score was 9. Median pre-treatment SUV was 8.0, and treatment delivered a median BED of 112.5 Gy. None of the patients experience local failure, and the median overall survival was 8.1 months (mean: 13.6 months). Finally, only one of these patients reported further disease progression (see page 16 lines 316-321).

153: Was the response done according to RECIST?

Reply 12: Yes, the RECIST criteria were used, and this phrase was added to the methods section, accordingly.

Changes to the text: Radiographic response was graded as complete, partial, stable, or progressive "according to the RECIST criteria" (see page 12 line 230)

182: in results section only 1 sentence covers the co-morbidity score, which holds a prominent place in your final conclusion. Please elaborate.

Reply 13: It was further added that 23 patients had a Charlson score ⩾ 9, and only 14 patients had a Charlson score < 6. Further, the Charlson score is thoroughly discussed in the *Endpoints* section, where 36-month overall survival estimates for Charlson score ⩾ 9 and Charlson score < 9 were also added.

Changes to the text: 23 patients had a Charlson score ⩾ 9, and only 14 patients had a Charlson score < 6 (see page 14 lines 270-271).
Also: 36-month overall survival was 0% for patients with Charlson score ⩾ 9, and it was 50.3% for patients with Charlson score < 9 (see page 16 lines 303-304).

186: versus 143; is a median of 60 Gy correct?

Reply 14: This is correct. The methods section describes that almost all treatments involved the delivery of 40-60 Gy to the tumors, and since the majority of the cases involved treatment with 60 Gy, the median dose was correctly reported as 60 Gy.

Changes to the text: None.

Minor comments:
61: nutrition. Do you mean weight loss?

Reply 15: In this setting, nutrition refers to nutritional status, as evidenced by values such as body mass index and serum albumin. This phrase has been clarified to read "nutritional status," as opposed to "nutrition."

Changes to the text: Guidelines for selection of patients with lung cancer for surgery...cardiovascular fitness, "nutritional status," and performance status... (see page 6 line 85).

63: the operations available. What do you mean? Please specify. Why mention SCLC since this entity is rarely treated with SBRT?

Reply 16: We agree that this sentence could be further elaborated. In light of the other comment to shorten the introduction, the section of this sentence in parentheses (including the comment about operations available and small cell lung cancer) was eliminated. "Operations available" was meant to refer to the range of surgical options available for resection (ranging from sublobar resection to larger lung resection with systematic lymph node dissection). This discussion is, however, beyond the scope of this paper. Choosing between these different surgical options is complex, with sublobar resections being an option for patients with low pulmonary reserve, but with the increased risk of local recurrence. On the other hand, approaches such as a carinal resection are technically challenging. Small cell lung cancer also presents a distinct pathological entity with different treatment guidelines, accordingly.

Changes to the text: The phrase in parentheses further describing "operability" was deleted (see page 6 lines 86-87).

61-65: It is unclear what is meant by high-risk patients. My suggestion would be to specify the factors described in 61-63 in negatively versus positively associated factors.

Reply 17: This section has been reworded to better specify this point. In the study cited, "high-risk" patients were defined as those with preoperative FEV_1 or preoperative D_{LCO} ≤50%. Some providers may use different thresholds (e.g. <35%) for these variables to define different levels of reduced baseline pulmonary function to influence surgical decision-making. These subtle distinctions are beyond the scope of this paper. Even so, the key point that while some "high-risk" patients may benefit from surgery, their outcomes are generally worse than patients with better baseline pulmonary function is important. Many such patients are included in the present analysis and are treated empirically with SBRT because the surgical outcomes may be somewhat less than optimal.

Changes to the text: Though "some patients with poor baseline pulmonary function" may still benefit from surgery…in "such" "high-risk" patients… (see page 6 line 87-89).

66: CT as an abbreviation is used here for the first time. Please specify the

complication, 5.7% refers to pneumothorax only or were there other complications as well?

Reply 18: CT was written clearly as computed tomography since this was first time of its use in the manuscript. The 5.7% risk includes other complications, as well, in addition to just pneumothorax. Specifically, in the paper cited, a major complication was defined as: "pneumothorax requiring intervention, hemothorax, air embolism, needle tract seeding, and death."

Changes to the text: Obtaining tissue biopsy···via "computed tomography (CT)" ···carries "approximately a 5.7% risk of a major complication" (see page 6 lines 91-92).

110: Please define the criterium "hypermetabolic"; which threshold value did you use?

Reply 19: Generally, SUV ≥ 2 was considered to be the threshold for a lesion being hypermetabolic. Rarely, a treatment might still be performed if SUV was mildly elevated with other concerning features for malignancy (e.g. continued growth on CT with an SUV of 1.8). In such scenarios, a shared decision-making model with the patient was employed to offer treatment depending on the specific clinical situation.

Changes to the text: "Generally, SUV ⩾ 2 was considered the threshold for a lesion being hypermetabolic; however, in rare cases, treatment might be undertaken if a lesion demonstrated other concerning features (e.g. continued growth on CT with an SUV of 1.8)" (see page 9 lines 161-164).

116: Please write a number as a word at the beginning of a sentence.

Reply 20: We appreciate this suggestion. We have implemented this change throughout the manuscript, where applicable.

Changes to the text: Various numbers were written as words at the starts of sentences throughout the manuscript.

161: predictive of prognostic factors?

Reply 21: We agree that this word choice is superior. The word "predictive" was replaced by "prognostic."

Changes to the text: …was conducted to consider "prognostic" factors for the… (see page 13 line 239).

170: same comment as for 110.

Reply 22: The definition of "hypermetabolic" in this study was previously discussed in reply 19.

Changes to the text: None.

233: same comment as for 110.

Reply 23: The definition of "hypermetabolic" in this study was previously discussed in reply 19.

Changes to the text: None.

Table 1 and 2:
The absolute and relative numbers could be combined in 1 column. The rate could be displayed as percentage e.g. Male 41 (46%) instead of using a rate with decimals.

Reply 24: We have implemented this change in both tables.

Changes to the text: Both tables were changed accordingly (see pages 37-40 lines 714-719).

General comments:
I would recommend the use of 1 decimal for age, DFS, OS and so on. The use of 2 decimals distracts the attention since this is not commonly used.

Reply 25: We appreciate this suggestion and have implemented this change throughout the manuscript.

Changes to the text: 2 decimal places were truncated to one throughout the

TRANSLATIONAL LUNG CANCER RESEARCH

manuscript.

Reviewer B
General comments:

The manuscript presents investigating an improved setup protocol for maintaining patient setup accuracy in conventional radiotherapy for lung cancer. Overall, current strategy would contribute reducing setup error with easy use and versatile protocol for a large number of clinical facilities. However, I would like to mention some major points;

1. Why did the couch adjust in the only AP direction? With simply consideration, systematic error would reduce with three direction adjusting (AP, SI and LR) rather than one direction adjusting. Please clarify this reason in the introduction section.
2. Although you investigate the setup and random error, did you consider the error in PTV margin? In addition, how did you decide PTV margin (median margin of 8 mm) in your facility? Please describe more detail.
3. The sentence of "the CTV was equal to the GTV or ITV" is inadequacy. In general, GTV, CTV and ITV have different volume. CTV is defined by added a margin for sub-clinical disease spread which cannot be fully imaged, and ITV is defined by considering intra-fractional target motion. Please describe more detail of target definition.

-minor comments-
Line 58-59 "the offline correction strategies may trigger a large random geometric deviation" Is it true? I think that offline correction may trigger a systematic error. Please check your result and cited references.

Line 68 Please show machine energy.

Line 79 "after planning" Is it correct? "before planning" or "after first simulation" ? Please check the sentence.

Line 82 "the superposition method" Did you use AAA method? Please describe more detail including its version.

Line 87 "prescribed for the PTV" Please clarify the prescribed point. Did you place reference point? or normalized at isocenter? Please describe more detail.

TRANSLATIONAL LUNG CANCER RESEARCH

Line 111 “my couch” and “My couch” are incorrect between this line and figure 1. Please correct.

Line150 “systematic and random errors in the AP direction are lower in the present study than in previous studies” Why the random error was lower than previous studies? Please describe more detail.

Line 183-184 “during beam delivery” Is it correct? I think 4-diminsional CBCT has been used to evaluate tumor motion before beam delivery not during beam delivery. Please check the sentence.

Line 253 “ ···in prostate cancer dcan offline···. ” The cited title is wrong. The title is “···in prostate cancer -can offline···.” Please check again carefully in reference section.

Reply 1: Upon discussion with the editorial office, it was determined that these comments were made regarding a different manuscript.

Changes to the text: None

AME Publishing Company