Peer Review File

Article information: http://dx.doi.org/10.21037/tlcr-20-627.

Reviewer A: Major Revision

Comment 1:

Authors reported a deep analysis about the challenges in the target volume definition in lung cancer (LC). From my point of view, this is a good article that analysed some critical issues in the definition of the gross tumour volume (GTV). Some of these issues are very known in the radiation oncologist community, while new frontiers are opening as MR-Linac and artificial intelligence. I would encourage authors to underline more deeply some aspects

Reply 1:

Thank you for the positive feedback about the article. We do acknowledge that the MR linac and artificial intelligence are two emerging fields that will definitely change the way we define the GTV in lung cancer in the future. These issues are addressed as indicated in the subsequent comments.

Comment 2:

It could be nice that for each paragraph, Authors underline not only pro but also cons for each imaging modality analysed. This could be very useful for readers in order to suggest potential new research fields

Reply 2:

The main limitation of using multimodality imaging is that very few centres have dedicated PET-CT and MRI scanner that can be used for radiotherapy planning. This results in images being acquired using a different position at different time periods. The position and size of the tumour might change in between scans. This results in mis-registration between the planning CT and the diagnostic scans leading to differences in interpretation between observers, as explained in section 4.1. Additional limitations for using FDG PET are included in section 4.3. A discussion on the emerging role of MR guided radiotherapy is also included in section 4.8 (refer to comment 3)

Change in manuscript: Section 4.3:

However, FDG PET-CT also has a number of limitations. PET has a low spatial resolution and can not detect very small nodules (<1cm). False negatives and positives may occur in diabetic patients with high blood glucose levels at the time of scanning. Increased FDG uptake is observed in many non-neoplastic lesions, granulation tissue (e.g. wound healing), infections and other inflammatory processes, eventually resulting in false negatives and false positives (41).

Comment 3:

Paragraph 2: this is a very important paragraph. What do you think about adaptive radiotherapy in NSCLC in order to decrease healthy tissue irradiation that could be considered as a major deviation during radiotherapy (RT)? Add some lines about this point.

Reply 3:

We do agree that the definition of the GTV on the planning CT only represents a snapshot in time of the tumour shape and position. In view of this, we introduced this issue in paragraph 2 and discussed some techniques that could be used to overcome this issue in section 4.7.

Changes in manuscript:

"Paragraph 2: Another important limitation is that the final GTV delineation represents a snapshot of the tumour shape and position in time. The tumour can change during the course of treatment as a result of changes in respiratory motion, tumour baseline shifts, regression and progression and anatomical changes caused by pleural effusion and infiltrative changes (1). Large safety margins are required to account for this uncertainty, potentially limiting dose escalation. Image-guided radiotherapy has, therefore, a crucial role in identifying these changes during treatment and various techniques have been proposed to adapt the treatment accordingly".

Comment 4

Paragraph 4.3 MRI in LC: this is one of the most relevant issue, specifically in the next future. In fact the use of MR-Linacs in the daily clinical practice is progressively increasing. I would encourage Authors to underline what are also the critical points in MR-Linac treatments in LC. In fact, during the treatment delivery, some commercial machine did not offer respiratory tracking during the IGRT verification. This is responsible for several movement artefacts and opens an important issue about how to manage the adaptive MR-Linac treatments (adapt to shape and adapt to position). In fact, the MR-Linac should support radiation oncologists to re-adapt daily the target volume and treatment, increasing the interobserver variations.

Reply 4

There are various adaptive strategies that could be used to adapt the treatment for variations in tumour position and shape both for CBCT and MRI. Although ART strategies have the potential to increase the accuracy of treatment delivery and hence reduce the dose to normal tissue, there are a number of challenges that need to be addressed for implementing these techniques in clinical practice for both CBCT and MRI guided radiotherapy. However, the superior image quality of MRI, together with the provision of functional information and real-time image acquisition without the use of ionising radiation will definitely open new doors for the development of adaptive strategies in radiotherapy. The advantages and limitations of adaptive strategies for both imaging techniques are now discussed in section 4.8.

Change in manuscript Section 4.8:

With the integration of cone-beam computed tomography (CBCT) and MRI on the linear accelerator, it is now possible to identify intrathoracic anatomical changes prior to treatment and adapt the treatment accordingly if necessary. During adaptive radiotherapy, the planning CT is first registered with the localisation image, and any variations in the tumour and OAR shape and position are assessed. This is then followed by the application of an adaptive strategy. These strategies can be divided into two categories, 'adapt-to-position' (ATP) and 'adapt-to-shape' (ATS)(2). For ATP, rigid image registration is used to assess and account for variations in the isocentre position only (for e.g. by adjusting the couch position). On the other hand, for ATS strategies, deformable image registration is used to transfer anatomic contours and dose between the CBCT and planning CT images. This is used to assess dose deviations caused by the intrathoracic tumour and anatomical changes, providing guidance to when the dose distribution must be reoptimised. In general, contour propagation is followed by contour editing, creating a new source of inter- and intra-observer variation that has not received much attention yet.

Intrathoracic tumour and anatomical changes have been reported in 72% of NSCLC (1) with about a third requiring adaptive therapy to ensure tumour coverage and reduce lung dose (3). Replanning to account for tumour shrinkage may reduce the dose to normal tissue and hence reducing toxicity. However, replanning needs to be balanced against the risk of missing microscopic disease. The LARTIA trial investigated the failure pattern in locally advanced-NSCLC patients with an adaptive approach (4). A re-planning was performed based on tumour regression seen on weekly CBCT scans performed during treatment in 50 out of 217 patients. A 6% marginal relapse and low incidence of acute pulmonary and oesophagal toxicity (2% and 4% respectively) were reported in this study. Several studies indicated that tumour volume change during treatment might be predictive for treatment outcome (5,6) and hence might improve current baseline prediction models for treatment outcome. However, these findings were not confirmed in the large study by Kwint et al. (7) that found no correlation between tumour volume changes and overall survival. Her findings indicate that ART after primary tumour regression might be safe, but this approach needs further validation in prospective trials. Functional tumour information from MRI and PET-CT may also have an important role in developing prediction outcome models. Furthermore, the implementation of ART techniques in routine clinical practice still remains challenging. Adaptive treatment changes can be performed offline between treatments, online immediately prior to treatment delivery, or in real-time during treatment. Online and real-time adaptations improve treatment delivery accuracy, potentially allowing for margin reduction (3). However, these come at the cost of increasing treatment time and may not be feasible for all tumours. On the other hand, the optimal time points and cutoff points for offline replanning are still not known and could be different for individual patients. Replanning is time consuming, and the accuracy of the dose evaluation depends on the accuracy of the deformable image registration and the accuracy of autosegmentation tools on CT, CBCT or MRI. The latter is currently limited for the definition of lung tumours (3).

The introduction of onboard MRI on the linac is opening new doors for adaptive radiotherapy in lung cancer. The MR linac allows for the acquisition of high-quality soft-tissue contrast images with functional information without using ionising radiation, allowing the oncologist to make daily treatment adaptation. Furthermore, MR-linacs now allow for cross-sectional beam-on imaging, making it possible to monitor tumour and organ at risk motion during treatment delivery without the need to use external surrogates or statistical respiratory models.

This, together with the ability to acquire images in the sagittal and coronal plane results in higher image quality with less binning artefact, and more realistic motion estimation as the uncertainty from an imperfect external-internal surrogate is eliminated. Moreover, it also facilitates the use of gating and tumour tracking techniques.

Nevertheless, there are a number of challenges that need to be addressed for the clinical implementation of the MR-linac (65). Patient movement can increase as a result of the prolonged treatment time and the claustrophobic environment of the MRI. Workflows and imaging sequences need to be developed for radiotherapy purposes. Software also needs to be developed to account for the lack of tissue density information required for dose calculations and the time consuming step of contour propagation, editing and QA should be optimised, for instance by introducing simultaneous remote review. Ultimately, the clinical and cost-effectiveness of this technique must be proven with well designed clinical trials.

Comment 5

What about artificial intelligence as a learning process in order to collect GTVs NSCLC contouring, with the aim to support progressively radiation oncologist in the outline process?

Reply 5

Autosegmentation tools based using machine and deep learning algorithms are now being developed for contouring of both tumours and OAR. These algorithms are showing promising results for various OAR as these they do not tend to vary significantly in shape and texture between patients. However, the development of these algorithms for lung tumours is more complex due to the large variation between patients. The lack of reliable gold standards makes it difficult to develop databases required to train and validate these algorithms. Another issue to consider is whether to use supervised or unsupervised training data sets. These issues are further explained in section 5.

On the other hand, while the introduction of artificial intelligence has the potential to automate the delineation process fully, overreliance on automation can introduce new errors. Artificial intelligence should ultimately assist and not replace human judgement. Our recommendation would be to develop new tools that could be used to assist the oncologist in reviewing the contours as suggested by the study of Hui et al. These issues are now discussed in more detail in section 8.

Change in manuscript: Section 5

The main barrier for clinical implementation of machine and deep learning algorithms is the availability of high-quality clinical contouring data for training. This data is often stored in secure servers across a number of hospitals that are not linked. Improvements in workflows and logistics would be required in order to securely link all the patient data required to develop contouring databases (8). An important question is whether to include all the oncologists' contours in the training database. Training of algorithms can be supervised whereby the algorithm learns from labelled datasets (i.e. good contours) or unsupervised whereby the algorithms tries to make sense of unlabelled data (i.e. providing contours that

have not been peer-reviewed) by independently extracting features and patterns from the images.

Training of algorithms using non-reviewed physician contours can introduce a bias by the particular physician's medical training, experience, goals, or misconceptions, eventually leading to an inaccurate segmentation (9,10). An extensive database is required to reduce the effects of major outliers, but this will also increase the computation time. This problem can be resolved through the use of supervised training data whereby only the contours that have been delineated using a specific protocol and peer-reviewed by experts are included in the database (8). Alternatively, only the contours from patients that had acceptable local control rates and toxicity could be used to develop the training database. The latter would automatically exclude cases whereby the tumour recurred as a result of a geographical miss or cases that had unacceptable toxicities due to an excessive inclusion of normal tissue. The limitation of this approach is that it still requires a manual intervention to label the data making it time consuming to develop the algorithm. Also, cases, where the PTV coverage is compromised due to proximity or OARs, may need to be excluded.

Change in manuscript: Section 8

Artificial intelligence could also be used to develop computer-assisted peer review software. Hui et al. (11) developed an algorithm that could be used to evaluate OARs in the thoracic region. In this study, the researchers simulated common delineation errors, including boundary deviations, missing slices, incorrect labelling, and craniocaudal over-extension for OARs in the thoracic region. The algorithm was able to detect 37% of the minor and 85% of the major errors. The reason for lack of precision in detecting minor errors was attributed to the fact that these errors were inconsistently judged by the reviewers. The use of this tool also improved the reviewers' error detection sensitivity from 61% to 68% for minor errors and from 78% to 87% for major error. The findings of these studies suggest that such tools could be used to assist the oncologists in reviewing contours, but they should not be used to replace human judgement. Over-reliance on the system might end up becoming counterproductive and actually reduce the ability of the reviewer to identify errors. Further research is required to develop similar algorithms for lung tumours.

Reference: Hui CB, Nourzadeh H, Watkins WT, Trifiletti DM, Alonso CE, Dutta SW, et al. Quality assurance tool for organ at risk delineation in radiation therapy using a parametric statistical approach. Med Phys. 2018;45(5):2089–96. doi: 10.1002/mp.12835

Reviewer B: Minor Revision

Comment 1

This paper provides a critical, exhaustive background for the guidelines on target volume definition in the radiotherapy treatment of lung cancer, available in the related literature, e.g., those recently published by the ESTRO ACROP group [ref. 33]. In fact, the translational ground of this subject is limited by the quality of the reference "gold standards", that is, a level of evidence consisting in experts' consensus, which in turn rely on mathematical modelling of clinical radiobiology parameters (that is, Tumor Control Probability – TCP, and Normal Tissue Complication Probability - TCP). In fact, comparative random trials evaluating clinical outcomes are lacking, to this regard. Very limited outcome data are available, except for a single abstract [ref. 24], cursorily and summarily reporting the negative impact of protocol violations on overall survival results, from a not specifically designed prospective random trial.

Reply 1:

> Identifying a specific link between target delineation errors and overall survival is not easy due to the large number of confounding factors that are involved. Radiotherapy protocol violations (not necessarily target delineation errors) has been linked to worse survival in the CONVERT and PROCLAIM lung radiotherapy trials and also in other sites as shown in the meta-analysis of Ohri et al. These references have now been included. Other studies highlighted in table 1 reviewed plans and found major target delineation errors in about 17% of cases.
>
> Change in manuscript section 3
>
> Protocol violations have been linked to worse survival in the CONVERT and PROCLAIM lung clinical trials (24,25) as well as other sites (26). Lack of experience, training and professional background has also been found to contribute to interobserver variation (19,23,27).
>
> The following references were included Brade AM, Wenz F, Koppe F, Lievens Y, San Antonio B, Iscoe NA, et al. Radiation Therapy Quality Assurance (RTQA) of Concurrent Chemoradiation Therapy for Locally Advanced Non-Small Cell Lung Cancer in the PROCLAIM Phase 3 Trial. Int J Radiat Oncol Biol Phys. 2018;101(4):927–34.
>
> Nitin Ohri, Xinglei Shen, Adam P. Dicker, Laura A. Doyle, Amy S. Harrison, Timothy N. Showalter. Radiotherapy Protocol Deviations and Clinical Outcomes: A Meta-analysis of Cooperative Group Clinical Trials. J Natl Cancer Inst. 2013;105(6):387–93.

Comment 2

The present radiotherapy procedures ensure high precision in radiation dose deposition, often beyond the definition of tumour margins achieved by morphologic and functional imaging. This can inherently cause topographical marginal missing, and also enhances the possibility of defective target coverage due to organ motion, of particular concern in lung cancer irradiation. The consequent risk of failure in local control may be limited by suitable, highly sophisticated tools. Thus, a rigorous comparative assessment of the most advanced procedures is warranted, intended to achieve the best delineation of Gross Tumor Volume (GTV) and, consequently, the appropriate identification of

Clinical Target Volume (CTV), Planning Treatment Volume (PTV), and Organs at Risk (OARs) contours.

These authors provide qualitative and quantitative analyses of the dedicated protocols, necessarily considering as surrogate endpoints the interobserver variations and protocol deviations, through an extensive and thoughtful review of the related literature. The evidence emerges that the performances of the treatment plan and delivery can be improved by highly effective technology resources, dedicated training, peer review, and multi-disciplinary approaches.

This reviewer appreciates the quality of this paper for its clarity, rigorous methodology, significance, and accuracy. In his opinion, the authors have properly omitted the reductionistic commonplace regarding the necessity of prospective trials, dedicated to the particular subject here dealt with. This approach, in fact, could hardly achieve significant results, using as primary endpoints the clinical outcomes, given the complexity of the variables to be taken in account, especially for patients undergoing combined drug and radiation therapy. Further, intentionally comparing more to less evolved technologic and procedural resources could be unethical. Rather, the opportunity may be indicated of large, multi-center "real world" data collections, prospectively sharing a common ontological status and advanced statistical methods of analysis, by the Radiation Oncology scientific Communities.

Reply 2

> We would like to thank the reviewer for the positive feedback on the article. We do agree that to date, it is still not possible to completely eliminate the interobserver variation in target volume definition; however, multiple interventions are required to optimise to reduce this variation. With reference to the comments made by reviewer 1, Artificial intelligence could have an important role in reducing this uncertainty and to automate this process. The main barrier for developing this technology is the availability of high-quality data that could be used to train and validate AI algorithms and also to facilitate the review of contours. This is now further discussed in section 6 and 8.

Comment 3

In conclusion, he recommends the publication of this paper, considering this last suggestion. More, he suggests to avoid the general statements regarding the "poor soft-tissue contrast" of CT, that doesn't apply for the aerated lung tissue, also considering that appropriate contrast-enhancement techniques and the present, high-resolution CT technology may obviate this limit in many cases.

> Reply 3: We do agree that CT provides an excellent visualisation between tumour and lung tissue, but we should also like to point out that CT has limits to discriminate tumour, soft-tissue and infiltrative changes. Contrast enhancement does improve soft-tissue contrast. However, lung tumours are often surrounded by soft-tissue that has a similar density, and with this will ultimately make it difficult to distinguish boundaries. High-resolution diagnostic CT scans can be used alongside treatment planning CT scans to assess interstitial lung disease and lymph node involvement. We have reworded the abstract and section 4.1 to make this more clear.
>
> Change in manuscript: Abstract

However, due to the limited contrast between tumour and non-malignant changes in the lung tissue, it can be difficult to distinguish the tumour boundaries on CT images leading to large interobserver variation and differences in interpretation.

Change in manuscript: Added Section 4.2

Intravenous iodine contrast can be used to improve the contrast between the tumour tissue and blood vessels. However, due to underlying co-morbidities, not all patients can tolerate intravenous contrast (36). Diagnostic high-resolution CT scan can be used alongside treatment planning CT scans to improve the assessment of interstitial lung disease and lymph node involvement (37).