Author’s response to reviews

Title: Computed Tomography Volumetry of esophageal cancer - The role of Semiautomatic assessment

Authors:

Yi-Hua Zhang (yi-hua.zhang@ki.se)
Michael Fischer (dr.michael.fischer@gmail.com)
Henrik Lehmann (henrik.lehmann@karolinska.se)
Åse Johnsson (ase.johnsson@vgregion.se)
Ioannis Rouvelas (Ioannis.Rouvelas@ki.se)
Gunnar Herlin (gunnar.herlin@karolinska.se)
Lars Lundell (lars.lundell@ki.se)
Torkel Brismar (torkel.brismar@gmail.com)

Version: 2 Date: 11 Sep 2018

Author’s response to reviews:

Dear Professor Marcus Carlsson,

We appreciate the effort by you and the reviewers. The thorough and detailed reviews have been of great value in improving the manuscript. We have responded to the comments on a point by point basis below. We hope that you now will find our manuscript suitable for publication in BMC Medical Imaging.

Sincerely,

Yi-Hua Zhang Division of Medical Imaging and Technology
Department of Clinical Science, Intervention and Technology
Karolinska University Hospital, Huddinge,
141 86 Stockholm Sweden
Phone: +46 73 949 48 84,
Fax: +46 8 711 48 40
Sophia Zackrisson (Reviewer 1): General comment:

This is a small study examining the value of a semiautomatic method for tumour volume measurements compared to manual in esophageal cancer. It seems like there is scarce literature on the subject, why the study may make some contribution to the existing literature, albeit the small sample size. However, I have difficulties to grasp the study since it is not clear to me which measures that were compared with each other and how the software works. The study needs to be reported in more detail and guide the reader better. In addition, there are some inconsistencies that have to be clarified. Please see my specific comments below.

Abstract Ok


Methods:

1. Why were only 23 patients included over 5 years? There must have been more patients in the institution during that time frame. Please explain.

We have clarified this in Methods. This study was a sub-study of a larger randomized multicenter randomized clinical trial. Details about inclusion and exclusion criteria, are now described under the heading Patients.

2. Were both genders included? Please specify how many men/women.

The included patient population has been detailed in Table 1 after the suggestion of Reviewer 2. Both genders were included.

3. Were both adenocarcinomas and squamous cell cancer included?

Histological typing of the tumors are detailed in Table 1.

4. How was the patient selection done? What were the prerequisites for a patient to be included? I would like to have a more detailed description of the patient selection process.

5. Was this study part of another study? In that case, please refer to that article Was it only newly diagnosed patients with a primary esophageal cancer (ICD code etc)?
6. How was the diagnosis confirmed?

7. Were patients with metastatic diseases included?

The section below the heading Patients has been updated clarifying that this study was a substudy of a larger randomized multicenter randomized clinical trial. Further details regarding the inclusion and exclusion criteria, are now described under the heading Patients. Details regarding confirmation of diagnosis are also included. Patients with metastatic disease were excluded.

8. "Spiral CT in the portal-venous phase..." This is stated in methods line 4. Then on line 12, same page, it is stated that the segmentation was performed in arterial and venous phase. Please clarify.

We have clarified that spiral CT of both arterial and portal-venous phase was acquired and used for segmentation.

9. Were there any restraints regarding the CT technology (technical parameters, slice thickness, dose, type of CT machine etc?). Please, develop this part more. The reader needs more information in order to be able to reproduce the study.

The details regarding the CT parameters have been clarified under the new heading “CT Imaging acquisition parameters”

10. I would like a more detailed description of the software used for the semiautomatic segmentation. Has it been used and described in another study? Please reference in that case.

We have added a reference (24) where a more recent version of this software was used for semiautomatic segmentation and compared with other similar software.

11. The measurements could be manually adjusted by the readers. Was it recorded how often this was the case? Were the readers allowed to look at other images or reconstructed planes to get better oriented etc? Please clarify

The number of manual corrections of the semiautomated segmentation was not recorded during the measurements. Only transaxial images were available to the readers. This has now been clarified.

12. The semiautomatic method was performed twice in both arterial and venous phase. It is not clear to me how this was done. Do you mean one measurement in each phase or
actually twice per phase? If the latter, which measurement was used, or was the mean used? Please clarify.

This has now been clarified in text. Each reviewer performed one measurement for each phase and method. Intra observer comparisons were done directly to the corresponding measurement in the other phase, while the same phase was used for interobserver variability. (see also response to comment 21)

13. Statistics: Ok, seems reasonable to use correlation, but have you considered using intra class correlation?

The Intra class correlation coefficients have been calculated and tested for significance.

14. Minor comment: the SPSS version 11 seems very outdated. Version 25 is now used.

All statistics have been recalculated using R 3.4.3

Results:

15. line 9: mean tumor volume was 46 ml. Please specify if this was the manual or semiautomatic method. Why not give the mean and range for both methods?

The presentation of tumor volume has been revised. Both manual and semiautomated volumes are now presented with range.

16. Data sharing: line 18. Maybe this is where the journal wants this information; but I think it would be better placed in methods under a separate headline.

This section has been moved to methods.

17. **Table 1: Here, the number of patients is 41? Why? 23 was stated before.

We have corrected the discrepancy. The table has also been replaced.

18. To get at better picture of the distribution of the two different measurements and of all patients, it would be helpful if this was illustrated by Bland-Altman plots.

Bland-Altman plots have been generated and are included as Figure 2 A-D
19. Maybe I am confused, but I do not understand how you actually compared the different measurements. Perhaps this means that you have to explain this better to the reader.

The comparison of the different measurements have been clarified. In addition to ICC, we also compared the absolute difference of each tumor measurement from the corresponding tumor mean measurement in order to determine the average percentage difference in volume.

20. There must be more comparisons to be made: 2 radiologist x 2 manual measurements x 2 semiautomatic measurements = 8?

We have merged the arterial and portovenous measurements which results in 2 radiologist doing a total of 46 measurements using each method, resulting in comparisons between consultant and resident radiologist using manual segmentation and between consultant and resident radiologist for semiautomatic segmentation.

21. Could you also report intraobserver agreement?

intraobserver measurement is now reported through the comparison of arterial and portovenous volume, as each tumor was only measured once for each contrast phase.

22. Was there a difference between the resident and the senior radiologist? Or how did you treat the data?

The agreement between the resident and the senior radiologist was compared using ICC and a significant difference of ICC was found. The difference of percentage in comparison the average volume the measured, the less difference of volume between the observers, the less variability induced by the method.

23. In any case, it has to be more clearly described in the results section, what group of measurements that was compared to what.

The Result section has been revised for increased clarity of which groups were compared with each other.

Discussion:

24. Please summarize the major findings of you study in the first paragraph of the discussion.

The findings of the study have now been summarized.
25. You state that you have excellent agreement on line 10. You did not report a measure for agreement. Further, you state that the measurements were independent from the level of experience. I cannot see that you actually reported that in the results.

ICC has now been included as a measure of agreement, the results show a significant increase of ICC when using semiautomatic segmentation.

Anders Sundin (Reviewer 2): General points:

Unfortunately, this manuscript is not well prepared. I get the impression that the authors have been in a hurry to submit the paper and this shows in the preparation of the manuscript. Many references are old. There is just one reference (17) from 2006 in the manuscript to support the key point of the study that the morphological tumour volume is important for the outcome of radio-chemotherapy. Are there additional up to date references to support this? The references on FDG-PET/CT are from 2007 and 2010 which is quite some time ago.

We have included more recent references (22, 35-37) where the segmentation of tumor volumes on CT images are used to delineate tumors for further analysis using textural analysis. We have also added a reference to a more up to date FDG-PET/CT study (21).

26. Current research is focusing on the metabolic tumour volume on FDG-PET/CT rather than the morphological tumour volume on CT and the metabolic response as crucial factors for the outcome and recently MRI including DWI has also come into focus in these regards. Why is this not included at least in the discussion?

More recent references have been included and a paragraph addressing current developments of DWI-MRI and FDG-PET/CT has been added to acknowledge this.

Detailed points:

Methods:

27. How were the patients selected? Which inclusion criteria were used? Consecutive patients?

Why 2007-2012? Why not more recent examinations?

This has been addressed through further details regarding inclusion and exclusion criteria.

28. "Arterial and venous phase" is mentioned on Page 6 line 12 and also on line 18.

We have revised this accordingly and removed one duplicate mention.
29. Page 7, line 9: What was/were the reason/s for the wide HU range 0-1000? Was there any part of the tumour that during contrast-enhancement measured more than 200HU?

Is this wide range necessary for the calculation? Please explain!

The rationale for the wide HU range was to ensure inclusion of all tumor tissue by leaving a large upper margin. This has been clarified by adding “in order to exclude air and include all esophageal tumor tissue.”

The lower limit is primarily to exclude air from being included into volume and the excessive upper limit is to ensure that all parts of tumor parenchyma are included.

30. More details needs to be supplied to understand how this was performed.

The method section has been expanded with more details regarding how the resulting groups were compared to each other.

Results

31. Page 8, Line 8-9. Was not a visible tumour on CT an inclusion criterium?

All tumors were visible on CT for all patients included in our study.

32. The full data needs to be shown

The full data is now available as Table 2.

33. Why not use the well-established Cohens kappa test to assess reader agreement?

Reviewer 1 (Sophia Zackrisson) suggested that ICC should be used to evaluate reader agreement. We decided to use ICC to assess reader agreement as it is more suitable for real number data.

34. Why not use the Bland-Altman plot to assess differences between the two methods to quantify tumour volume?

Bland-Altman plots are now provided as Figure 2 A-D.

35. Was tumour histopathology (adenocarcinoma vs squamous cell carcinoma) a factor that mattered in the precision of the volume assessment?
It did affect mean difference percentage, with a significantly lower difference for SCC in comparison to AC, but no significant difference in ICC was observed. The result section has been updated to describe these findings.

Discussion

36. What has been performed in this area are using MRI? Including DWI?

Further more recent studies including DWI-MRI has been included and acknowledged in the discussion.

Table 1.

37. According to the Materials & Methods section there were 23 patients. Now in the table 41?

According to the Materials & Methods section transverse 2.5 mm section were reconstructed/reformatted. There is no need to show that there were 0 patients with 3, 5, 7, 8 mm sections and no need to include this info in the table since it has already been mentioned. At bottom of the table what does "Unknown 3 1" mean? With as few as 23 patients there should be a table for all patients (23 rows) showing the measured volumes for both readers and both methods (4 columns). Exchange Table 1 for such a table and for each patient include the information on patient characteristics (at least age, sex), tumour histopathology, tumour stage and tumour localization in this new table.

Table 1 has been corrected and rearranged as suggested. The results from volume measurements have been written out in a separate table (Table 2).

Figure legend to figure 1.

38. *** "Axial" should be "transverse" or "transaxial"

This has been revised.

39. Which delineation has been used (manual or semiautomatic? Is it not interesting to include a figure 1B the show result for the other delineation method in the same patient!?

The figure shows manual segmentation. We are unable to add a figure for the semiautomatic segmentation as we no longer have access to the same software.

Figure legend to figure 2.
40. What is the indicated in the bars?

The bars show 95% Confidence interval. This has been clarified in the figure legend.