Author’s response to reviews

Title: Semi-Automated Analysis of Standard Uptake Values in Serial PET/CT Studies in Patients with Lung Cancer and Lymphoma

Authors:

John Ly (john.ly@skane.se)
Sabine Garpered (sabine.garpered@skane.se)
Peter Höglund (peter.hoglund@skane.se)
Eskil Jönsson (jonsson.eskil@gmail.com)
Sven Valind (sven.valind@skane.se)
Lars Edenbrandt (lars.edenbrandt@med.lu.se)
Per Wollmer (per.wollmer@med.lu.se)

Version: 2 Date: 28 January 2012

Author’s response to reviews: see over
Dear editorial board of BMC Medical Imaging,

Please find enclosed the revised manuscript: Semi-Automated Analysis of Standard Uptake Values in Serial PET/CT Studies in Patients with Lung Cancer and Lymphoma, by John Ly, Sabine Garpered, Peter Höglund, Eskil Jönsson, Sven Valind, Lars Edenbrandt and Per Wollmer to be submitted as an new article to Medical Imaging.

In this revision, we have addressed all concerns made by the reviewers. The most important change is the change of statistical analysis to Bland-Altman and we believe our results will be more readily understood and correct by this. We have also improved the introduction to provide a more accurate description of current research in the field of PET/CT quantification.

We still believe that our findings will be of interest to the readers of Medical Imaging because they bring evidence that semi-automatic quantification is timesaving and accurate as manual quantification of maximum standard uptake values.

We hope that the editorial board and the reviewers will agree on the interest of this study.

Sincerely yours,
John Ly.
Response to Ronald Boellard

1. Introduction: the authors state that semi-automated quantification studies are still limited...I do not agree. There are a lot of papers dealing with automated PET tumor delineation and several papers in image registration. I suggest the authors perform a careful search in clinical literature databases. Some hints, look for “Schaefers, Nestle, Black, Lee and Geets, Hatt and Visvikis, Cheebsumon and Boellaard, Van Velden etc etc”...there are many papers on tumor delineation. Likewise there are several reports on the alignment of serial PET/CT scans.

The wording was unfortunately faulty in this paragraph. We meant to specify computer-assisted quantification of SUV in serial examinations. But we were also wrong about lack of tumour delineation papers. The paragraph has been re-written to reflect the general status of quantification research and computer-assisted research.

2. SUVmax are compared with and without image registration functionality. However, SUVmax are calculated on different platform. According to the authors one of these platforms performs a smoothing of the images, while the other does not. Therefore, results cannot be compared directly, i.e. the reader cannot really tell if there are no differences in SUVmax based response data due to use of the image registration procedure or not.

We are not entirely sure whether PHILIPS platform performs smoothing of the images. In the original manuscript we were suggesting this as an explanation to the systematically higher SUVmax values on the PHILIPS platform but we’re not sure. In order to bypass this problem we’ve opted to use Bland-Altman analysis instead.

3. There is very limited information on the image registration method (which type of algorithm, rigid or deformable, cost function such as mutual information or correlation etc etc).

We have added additional information regarding image registration method in the revision.

4. Are the images transform or is the VOI repositioned between serial scans ?

The VOI is repositioned. We’ve added this information to the manuscript.

5. Although the automated method enhances overall processing time, I wonder if the improvement is really clinically relevant. How long does the processing and image registration take ? Usually image interpretation takes more time than doing the SUV read and changing from 1 min to 0.15 min is not really that important. Yet, the tools might be quite convenient to use. It would be more interesting to know if the tool can help to identify the same lesion in serial scans...any benefits there ?

The time to register two examinations take roughly 40 seconds but it’s heavily dependent on programming platform and may be reduced further. We have been experiencing an increased demand from our clinicians regarding long axis and SUVmax quantification of multiple reference tumours in serial examinations. In lymphoma or metastatic disease the number of lesions of interest can easily
become plenty, this is when the tool can really support the reader. The tool automatically identifies the same lesion in the next examination when the original tumour has been selected in the first examination, however when the lesions are in clusters we do not recommend using the tool.
Response to Irene Burger

1) Abstract:
Include in the methods part that 3 readers were analyzing the data visually and semi-automatically, and were compared with each other for lesion characterization with SUVmax and that time for analysis. The sentences: Patients with abnormal lesions showing sharp contrast to surrounding areas and with no formation of a large conglomerate mass were selected. However does not seem to be essential for the abstract.

This has been revised.

2) Introduction:
Might consider including Ref: Fox et al. Practical Approach for Comparative Analysis of Multilesion Molecular Imaging Using a Semi-automated Program for PET/CT, J Nucl Med 2011 52:1727-1732
Using a similar approach in prostate cancer.

This reference has now been added to the manuscript.

3) Result: As discretionary suggestion:
Emerging results show that, although the interobserver reproducibility is very high for SUVmax the repeatability of SUVmax is very low due to statistical variance due to the variability in nuclear decay. See: Repeatability of SUV measurements in serial PET. Schwartz et al. Med Phys. 2011 May;38(5):2629-38.
It therefore would be of great interest if your semi-automatic analysis also results in an as good or might even better reproducibility for averaged SUV than manual reading (SUVmean, or SUV hottest 10 or 20 pixels). This however is looking into another topic and could be considered in a second study using your new segmentation tool.

Very interesting idea. The tool does indeed segment the entire tumour volume and extracting SUVmean or SUV hottest 10 or 20 pixels can certainly be added to the tool.

4) Discussion:
a) The choice of performing the segmentation in the PET is because the patient material had both lowdose and diagnostic CT. Low-dose CT provides a lower resolution of the CT scan, which would most likely result in poorer segmentation quality.
This part is misleading: in an unenhanced CT segmentation of a lesion (unless lung) is inaccurate. Since Lymphoma was included and this was most likely not in the lung – CT-segmentation is not really an option and the Low versus High dose aspect is of secondary importance. Furthermore by being able to segment in PET this method can be used in a wider range of disease (hepatic metastasis...). This should be emphasized instead.

We have made re-written this part in the revision.

b) SUVmax was chosen over SUVmean due to clinical praxis at our hospital when assessing PET/CT examinations.
Since SUVmax is still the most commonly used value, this is certainly reasonable. However, SUVmax seems to be overestimated since it’s obviously in the same scan always the same value – however it seems to be lacking a reasonable reproducibility.

While SUVmax seem to lack reasonable reproducibility, only rare clinical settings would need a secondary PET/CT scan within a short period of time. The second examinations in the study were either post treatment or after a few months expectation.

c) Last line on page 10) there is an issue with reference 5 – please check.

This has been corrected.

5) Conclusion:
Good agreement was shown in absolute SUVmax measurements and but there was a significant difference in the mean difference in SUVmax between studies when comparing both methods.
Needs to be rewritten: There was no significant difference in SUVmax when comparing both methods?
Also include the significant time reduction into the conclusion with the semi-automatic method.

We have chosen to use a Bland-Altman analysis in our revision, and therefore the conclusion has also been re-written. We have now also included the significant time reduction in our conclusion.

6) Minor essential revision: Figure1 is the flow chart in the text. Figure 2 (regression analysis is wrongly entitled Figure 1)

This has been corrected.
Response to Eric Laffon

Abstract: - methods: indicate the number of readers in this section. - results: statistical analysis should be totally reconsidered (see below).

This has been corrected.

Introduction: - The second paragraph could be developed.

This paragraph has been completely re-written.

Materials and Methods
- Manual method: when comparing two PET examinations, was SUVmax always found in the same slice? Indeed, scanning the whole lesion in each examination often shows that the slice involving SUVmax can be different (even for 2 scans successively performed).

There seems to be a misunderstanding here. Tumours are segmented as a 3D volume in the PET image in both examinations after image registration. Thereafter the SUVmax of the whole volume segmentation is calculated. Changes to the manuscript have been made to clarify this.

- Semi-automatic method: the way this method is implemented introduces a bias in the study, for the above mentioned reason. In other words, the automatic alignment does not take into consideration that SUVmax could be found in a different (contiguous) slice than that marked in the first examination. (Note that a semi-automatic volumic assessment could avoid such a bias.)

Please see above.

- Statistical analysis: should be totally reconsidered by using the method of Bland and Altman, because linear regression is not sufficient to assess the agreement between 2 methods. (Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1:307-310.)

We have applied a Bland-Altman analysis on our material and omitted linear regression. Please see changes.

Results
- Linear regression analysis: this reviewer is not sure that there is no confusion between "intercept" and "slope" (please give additional details, such as equations). However, as mentioned above, the analysis is not appropriate.

Please see above.

- Range of SUVmax values: the lowest SUVmax value (2.4) is too low. A cut-off of 2.5 is usually used. Means should also be given.

The 2.4 value was obtained on the second examination post radiologic treatment. The lesion was still considered pathological. We do not agree that lower SUVmax
values should be omitted from this study if the lesions have been assessed as pathological.

- Table 1:

  . This reviewer does not understand why the column "manual" shows different figures between readers. Indeed, manual measurements should mandatorily provide similar results! (as much as "One of the readers marked the 26 lesions in screenshots ...to secure that all three readers were measuring the same lesions")

  . Last column: what is "p" related to? Moreover, the second reader found a significant difference between the two kinds of assessment that is not clearly emphasized in the results.

  . "The differences in manual measurements are systematically lower than the semi-automated ones". Why don’t you use a simple sign test to ascertain it?

This table has been removed because it was based off results from linear regression which has now been replaced by a Bland-Altman analysis.

  . Give unit of SUVmax.

Examinations acquired by our PHILIPS PET/CT use counts and a private DICOM factor to calculate SUV and has therefore no unit and thus deviate from other vendors which use bqml.

Discussion and Conclusion
- The discussion is not structured because the results are unclear (or clearly understood).

We have made changes to the discussion after including Bland-Altman analysis.

- The first sentence of the conclusion summarizes the deficiencies of the paper: "... but there was a significant difference in the mean difference in SUVmax between studies when comparing both methods". This point is neither clearly presented in the results section nor discussed, whereas it is the main interest of the paper.

This part of the conclusion has been re-written due to the change from linear regression to Bland-Altman analysis.

- Minor Essential Revisions:
  Title and Manuscript

  - Please choose between "semi-automated" and "semi-automatic" and keep on your choice throughout the manuscript.

  “Semi-automatic” has replaced “semi-automated” in the entire manuscript.

Materials and Methods - An estimate of the effective dose (mSv) might be given for both PET and CT.
This has been added.

Results
- I suppose that Figure 1 is flowchart and Figure 2 is erroneously labelled "Figure 1".

This has been corrected.

- Second paragraph: "almost-perfect reproducibility ..." perfect is awkward.

Interpretation of ICC was categorized according to Landis and Koch which has been used in other similar papers. According to Fleiss’ categorization our results would have been categorized as "excellent” which may be even more awkward.