Author’s response to reviews

Title: The diagnostic performance of $^{18}$F-FAMT PET and $^{18}$F-FDG PET for malignancy detection: A meta-analysis

Authors:

Arifudin Achmad (arifudinachmad@gmail.com)
Anu Bhattarai (bhattaraianu28@gmail.com)
Ryan Yudistiro (ryannuclear@gmail.com)
Yusri Heryanto (yusri_dh@gmail.com)
Tetsuya Higuchi (tetsuyah92md@gmail.com)
Yoshito Tsushima (yoshitotsushima@gunma-u.ac.jp)

Version: 1 Date: 16 Sep 2017

Author’s response to reviews:

COVER LETTER – MAJOR REVISION

General Response to Editors and Reviewers:

Dear the respected BMC Medical Imaging Editors and reviewers,

Thank you for reviewing this manuscript. Your comments are valuable and had definitely improved the quality of this manuscript. The authors had addressed each of the comments.

There is some important data we want to add to the manuscript, based on a suggestion from a statistical expert. This addition will not change the conclusion. Instead, it will add statistical evidence. Through a further statistical analysis (meta-regression for bivariate method) suggested, we obtained p values from hypothetical tests for average sensitivity and average specificity of each radiotracer in each meta-analysis (by visual assessment and by diagnostic cut-off value). We believe that this additional p values will give a clearer result for the readers.

The additional data we want to add are:

In Table 3, between the columns of $^{18}$F-FAMT and $^{18}$F-FDG,

- Based on the visual assessment, p values of:
  
  average sensitivity: 0.181, average specificity: 0.207
- Based on the diagnostic cut-off, p values of:

  average sensitivity: 0.542, average specificity: 0.009

Accordingly, we added in the Results section, P9 L28:

"There was no significant difference in average sensitivity and specificity between $^{18}\text{F}-\text{FAMT}$ and $^{18}\text{F}-\text{FDG}$ based on visual assessment (p = 0.181 and 0.207, respectively). However, $^{18}\text{F}-\text{FAMT}$ was significantly more specific than $^{18}\text{F}-\text{FDG}$ (p < 0.01) based on diagnostic cut-off values."

Accordingly, in the Abstract, P1 L47:

"While the average sensitivity and specificity between $^{18}\text{F}-\text{FAMT}$ and $^{18}\text{F}-\text{FDG}$ based on visual assessment were similar, $^{18}\text{F}-\text{FAMT}$ was significantly more specific than $^{18}\text{F}-\text{FDG}$ (p < 0.01) based on diagnostic cut-off values."

We also add one more co-author (YDH) regarding this valuable additional data.

Best Regards,

Arifudin Achmad

Response to Reviewers (Q: Question, R: Response)

Reviewer 1 (Dr. Monica Shooken):

Q1. What was the reasoning behind the use of Cochrane Collaboration over other methods? A rationale for using these recommendations should be described in some detail.

R1. Meta-analysis of data from studies of diagnostic test accuracy (DTA review) is more complicated than most other forms of meta-analysis (principally due to the paired nature of the main outcome measures sensitivity/specificity). Moreover, the complexity of DTA reviews is not restricted to statistical methods. Meta-analysis of diagnostic test accuracy studies is a particular type of meta-analysis which only developed within the last two decades (Irwing L, et al. J Clin Epidemiol 1995; 48(1):119-30); and its statistical method is still continuously developing.

The reason behind the use of Cochrane recommendation is that almost all of the statisticians and epidemiologist who developed the early mathematical method (Les Irwig, Paul Glasziou) until they who involved in the development of the current standard for the methodological framework of DTA reviews, including STARD and QUADAS tools (Jon Deeks, Penny Whiting, Mariska Leeflang, Johannes Reitsma, and many more) are all included in the consensus and the protocol project at the Cochrane (Leeflang M., et al. Syst Rev 2013; 2:82).
The Cochrane Collaboration provided a detailed protocol in their Handbook for DTA review: http://methods.cochrane.org/sdt/handbook-dta-reviews and step-by-step training in their website: http://training.cochrane.org/path/diagnostic-test-accuracy-dta-reviews-pathway. Other ‘method,’ for example, DTA review manual from The Joanna Briggs Institute (JBI), basically recommends the same statistical analysis methods:


Q2. For the extraction method, it would have been appropriate to use the last author's name along with the first author's name as the last author is generally the corresponding author with most expertise.

R2. Thank you for your suggestion. However, such approach is not mentioned anywhere in Cochrane nor JBI review manual. If this is essential information and recommended by the guidelines, we will find last author names also referred to in many systematic review/meta-analysis report.

Q3. Line 6-8: The sentence "Overall, the methodological…..was good" should be removed or edited to include objectivity such as significance level.

R3. Revised to: “Overall, the included studies have a low risk of bias with good methodological quality based on QUADAS tool.”

The detailed information of this sentence is available in Results (Study Eligibility, Quality, and Risk of Bias) P7 L32: “All were prospective studies of good quality (QUADAS Scores > 10) with at least involving 19 patients (patient number range: 19–74) and 21 lesions (lesion number range: 21–75). Overall, the nine eligible studies have a low risk of bias, except in blinding from the index test results.”

Q4. The conclusion statement is not fully supported by the results.

R4. The main results are summarized in Table 2, Table 3 and Figure 3.

In Table 2 (univariate analysis) and Table 3 (bivariate analysis), diagnostic odd ratios (DOR) and area under HSROC curve (AUC) of $^{18}$F-FAMT were higher (and larger) than that of $^{18}$F--FDG in meta-analyses either based on the visual assessment or diagnostic cut-off value. However, one might not be easily satisfied with single diagnostic value like DOR or AUC, but rather prefer separated reporting of pooled sensitivity and pooled sensitivity.

We recently performed a hypothetical test to evaluate the difference in average sensitivity and specificity of each diagnostic method (visual assessment or diagnostic cut-off value). The test results were added in Table 3 (as mentioned in general response above). There is no difference between average sensitivity of $^{18}$F-FAMT and $^{18}$F--FDG either based on visual assessment (p = 0.181) or based on diagnostic cut-off (p = 0.542). Based on the visual
assessment, the average specificity (or 1 - false positive rate) was similar (p = 0.207), while based on the diagnostic cut-off, $^{18}$F-FAMT was significantly more specific than $^{18}$F-FDG (p = 0.009).

A separated reporting of average sensitivity and specificity is recommended when studies report a common threshold for a positive result. A HSROC curve is recommended when studies report several different thresholds. In this meta-analysis, included studies reported different thresholds (cut-off values) for both $^{18}$F-FDG and $^{18}$F-FAMT. However, the Spearman correlation test (in Results, Descriptive statistics) showed that threshold effects might influence meta-analyses based on visual assessment, while threshold effects were not evident in meta-analyses based on diagnostic cut-off values. Therefore, we described both separate reporting of average sensitivity and specificity, and also HSROC curve.

In Figure 3 (A and B), sensitivity and specificity are represented by the position of summary operating points (SOP) within the HSROC graph. The similar height of the SOP in the Y-axis (sensitivity) showed that $^{18}$F-FAMT and $^{18}$F-FDG have similar sensitivity. In case of specificity, even though the X-axis distance between two SOPs was larger in Figure 3 A (based on visual assessment), one should note that the variability of the studies was also larger (represented by the wide confidence regions with substantial overlap). With less variability and small overlap in the confidence regions, the X-axis distance between $^{18}$F-FAMT and $^{18}$F-FDG SOP in Figure 3 B, adequate to demonstrate that $^{18}$F-FAMT was more specific than $^{18}$F-FDG.

Since both sensitivity and specificity played a major role in malignancy detection, we concluded based on the current finding that:

1. $^{18}$F-FAMT is more specific for malignancy than $^{18}$F-FDG while their sensitivity was comparable.

2. $^{18}$F-FAMT is equal to $^{18}$F-FDG in diagnostic performance for malignancy detection in several cancer types.

Q5. A major concern is that the dynamic data with FAMT especially at the early 5-10 minute time point is not available. This makes it hard to correlate with the biology of the tracer uptake as the washout of amino acid-based tracers occurs at early time points.

R5. Yes, unfortunately, clinical dynamic PET study with $^{18}$F-FAMT is currently lacking.

However, one animal PET study showed that the tumor-to-muscle contrast ratio of $^{18}$F-FAMT may still be acceptable within a period from 20 min (highest) to 60 minutes post-injection (Yamaguchi A, et al. EJNMMI Res 2015;5:29). We added this portion in Discussion as well as in reference (#38):

A dynamic $^{18}$F-FAMT PET study in animal tumor model showed that tumor-to-muscle uptake ratio is the highest at 20 min and still high enough at 60 min [38]. However clinical dynamic PET studies are necessary to obtain the optimal scan time.
Q6. Another major limitation of the study is that all the authors are part of the same institution where $^{18}$F-FAMT was produced. All six studies included in the analysis are also part of the same single institution.

R6. Yes, this is our limitation.

Here is our detailed explanation: All studies are generally from Gunma University. $^{18}$F-FAMT is produced by our department (Department of Diagnostic Radiology and Nuclear Medicine, previously: Department of Nuclear Medicine). However, any department in Gunma University/Gunma University Hospital which is in collaboration with us can use $^{18}$F-FAMT. From the six included study, the study of Dr. Inoue and Dr. Tian are the only works where the first authors belonged to our department (at the time their paper was published. Now they are no longer part of our department). Dr. Watanabe and Dr. Suzuki were from Department of Orthopedic Surgery, Dr. Miyakubo was from Department of Stomatology and Maxillofacial Surgery, and Dr. Kaira was from Department of Medicine and Molecular Science. More importantly, all the first authors of the included studies have no influence nor contact with the main authors of the current meta-analysis at any time during this study conducted (AA, AB, RY and YDH).

Dr. Tsushima (YT) and Dr. Higuchi (TH) were the 3rd and 4th author in Dr. Miyakubo’s study. Dr. Higuchi was the 5th author in Dr. Kaira’s study. However, in the current study, they were not involved at any time in the study selection, data extraction, meta-analysis nor the result interpretation. They indeed reviewed the final draft and permitting this meta-analysis to be published, but not directly involved.

Q7. To enhance the value of using FAMT tracer for certain cancer types, authors should tabulate the tumor type and then compare with FDG. Discussion as written does not provide a "trend" in tumor types or location favoring FAMT over FDG. While there is some data supporting this, this is not summarized in the discussion.

R7. Thank you for your suggestion.

In Table 1, we detailed the tumor types evaluated in the nine studies to describe the distribution of the lesion involved in this meta-analysis. For such “trend” in tumor types or location, nine studies may not be representative enough. There were 30 clinical studies that can be summarized to generate such tabulation. However, this is beyond the scope of this meta-analysis.

Actually, we prepared a table describing all $^{18}$F-FAMT clinical studies where the unique feature of $^{18}$F-FAMT in each study (in comparison with $^{18}$F-FDG) is briefly described. This table is included in our comprehensive literature review, which, unfortunately, BMC Medical Imaging has no rubric for it.

Reviewer 2 (Dr. Ali Azhdarinia):
Q1. Since the studies evaluated cover a period of 11 years, were there any changes in the synthesis of $^{18}$F-FAMT?

R1. No. The current $^{18}$F-FAMT synthesis method has not changed since the first used for animal study in 1997. The inefficient radiosynthesis is the main challenge for $^{18}$F-FAMT, and the recently booming ‘late-stage fluorination’ method has been considered. However, this new method has not experimented for $^{18}$F-FAMT yet.

Q2. Were the imaging parameters for the acquisitions similar for each tracer in trials (i.e., time post-injection)?

R2. Yes, similar. The injection dose is described in Table 1. Scan time post-injection is also similar (40 minutes in all six studies except in Dr. Kaira’s study, 60 minutes). The $^{18}$F-FAMT dynamic PET study in animal demonstrated that the tumor-to-background contrast is acceptable within the post-injection time of 20 min to 60 minutes (Yamaguchi A, et al. EJNMMI Res 2015;5:29).

Q3. Quality of Figures 1-3 needs to be higher.

R3. The actual image resolution of the submitted figures is already high and satisfied the quality requirement as mentioned in the author guidelines: Figure 1 (1200 dpi), Figure 2 (1200 dpi) and Figure 3 (600 dpi). The image quality may be reduced during automatic PDF generation for review purpose.

Q4. There are several issues with grammatical errors which need to be corrected.

R4. Revised and consulted to a native English speaker.