Author's response to reviews

Title: Diagnostic value of Diffusion-Weighted MR imaging in thyroid disease: application in differentiating benign from malignant disease

Authors:

yingwei wu (wuyw0103@hotmail.com)
xiuhui yue (wuyw0103@gmail.com)
weiwen shen (sww_281@sina.com)
Yushan Du (laifa1015@163.com)
Xiaofeng Tao (cjr.taoxiaofeng@vip.163.com)
Ying Yuan (yuany83@163.com)
Cheuk Ying Tang (cheukying.tang@mssm.edu)

Version: 3 Date: 10 January 2013

Author's response to reviews: see over
Dear Editor,

Thank you and the reviewers so much for taking the time to review our manuscript. We have made the appropriate changes as recommended by the reviewers and have also tried to address the comments to the best of our ability. Changes have been made to the text and figures. We have also added the ADC table as recommended by the reviewers. Below are individual answers to the questions from the reviewers and where appropriate we have made the changes in the text already.

Thank you,
Reviewer's report 1
Title: Diagnostic value of Diffusion-Weighted MR imaging in thyroid disease: application in differentiating benign from malignant disease
Version: 2 Date: 4 December 2012
Reviewer: hakan mutlu

Reviewer's report:
Abstract
#Background and Purpose#
...... to differentiate malignancy of thyroid lesions should be changed to differentiation of benign from malignant thyroid lesions. Done

The area of ROIs were high so it affects ADC values.
We appreciate the concern of ROI size. We also understand that there is no ideal ROI size since lesions vary in size as well as composition. We chose the ROI size to minimize noise and at the same time consistency.

FSE EPI may cause less artifacts in thyroid disease. But absolute ADC can be used as alternative method.
The ADC we calculated was based on the formula: $\text{ADC} = -\ln(S(b)/S(0))/b$ for the different $b$-values.

Limitations of this study.
1. First, small number of the malignant thyrod lesions limits the statistical power.
We acknowledge this shortcoming in the discussion. As this data was already acquired over an extended period of time and the pathology results (a unique aspect of this study) was not always available for every patient.

2. Thyroid lesions less than 1cm were excluded from the study. Improvement in the software and using FSE EPI of diffusion-weighted MR imaging may help in the detection of smaller lesions.
We agree with this, but the current FSE software is still not perfect in terms of geometric distortion and sensitivity to patient movement to warrant a reliable measure of regions smaller than 1cm.

Level of interest: An article of importance in its field
Quality of written English: Acceptable
Statistical review: No, the manuscript does not need to be seen by a statistician.
Declaration of competing interests:
I declare that I have no competing interests.
Reviewer's report 2
Title: Diagnostic value of Diffusion-Weighted MR imaging in thyroid disease: application in differentiating benign from malignant disease

Version: 2 Date: 5 December 2012
Reviewer: Frederik De Keyzer

Reviewer's report:

Discretionary Revisions:
1/ Background, last paragraph: “we investigated the characteristics of thyroid diseases”. It might be better to reformulate this, as you only look at the benign or malignant nature of the lesions.

Thank you for the comment. We have rephrased the sentence.

2/ Methods, MRI protocol, last paragraph: “total scanning time was about 10 minutes”. This is incredibly short for Ax T1, Ax T2, DWI and Cor T2. How many slices did you acquire to have such a short scan time?

Sorry for the confusion, the time indicated was for DWI only. We have now made it clear in the text.

3/ Discussion, MR technique: Please explain what type of modification you used in your SE-EPI sequence that would further reduce artifacts and improve quality. As far as I know, the study by Razek et al, and probably also by Wang, used the SE-EPI sequence, not just an EPI, as this is the only one routinely implemented on clinical MR scanners for the last 10 years.

This custom sequence was implemented on our 3T Achieva scanner and is based on ref#13 by Murtz et. al. 2007.

4/ Discussion, page 14: The entire paragraph on FNAB does not really add anything useful to the paper, as the reference standard was the histological results, and no accuracy, sensitivity, or specificity was provided on the FNAB in correlation to the reference standard; in my opinion, it can be left out.

We understand your concern on this. The reason we included it was to emphasize the uniqueness of our study which had access to pathology data. There have been other concerns that a similar study with FNAB as comparison may not be accurate.

5/ Figures 1 and 2: It would be nice to show the performed delineations on the images.

We have now shown figs with ROIs placed on lesions.

Minor Essential Revisions:
6/ Abstract, Materials and Methods: change ‘42.43 years’ to ‘42.4 years’ as it is used in the Results (consistent rounding).

Done

7/ Background, first paragraph: Please provide references for the statements “FNAB results may mimic some other diseases”, and “US is the most common … but there are still no reliable criteria for distinguishing benign from malignant
lesions.”
We have now included some references. The point was not to say that FNAB is inaccurate, it is still the gold standard with close to 94% accuracy, but it is not perfect.

8/ Methods, Patients selection, first paragraph: “... or ultrasonography determined thyroid were prospectively ...” should be “... or ultrasonography determined thyroid lesions were prospectively...”. Corrected now.

9/ Methods, Image post processing: Please state whether delineations on the different ADC maps were performed on each ADC map separately, or were they delineated on one, and then copied to the others?
Copied onto the others, now indicated in the text.

10/ Methods, Image post processing: “experience” should be “experienced”.
Done

11/ Results, first paragraph: “14 had lesions on both lobes”, what happened with the lesions? Where they examined as independent lesions, or did you average the values for the bilateral lesions? Please specify.
For those with lesions on both lobes, one was randomly chosen (matched with pathology) for ADC measurements. Now indicated in the text.

12/ Results, Parameters, last paragraph: ‘obatined’ should be ‘obtained’.
Done.

13/ Discussion, MR technique, third paragraph: “higher b values also leads to increased signal attention and usually required more averages to compensate for the SNR”. Please specify in the Methods section what averages were used for each b-value, as this affects your ADC calculation.
6 Averages were obtained uniformly for all b-values.

14/ Discussion, MR technique, last paragraph: “higher b-values also produce more susceptibility distortions and could increase the noise ...”. In itself, higher b-values do not produce more susceptibility distortions, although higher b-values usually lead to higher echo times, which in turn can cause more distortion.
Yes, we understand that. Also note that higher b-values also increase eddy current related distortions that are different depending on the gradient directions leading to increased noise.

15/ Author contributions: “paients” should be “patients”, “mesurement” should be “measurement” (twice).
Done

16/ References: Reference 12 and 18 are identical.
Removed & consolidated.
Major Compulsory Revisions:

17/ Throughout the manuscript, the definition of ADC calculation is misleading, as is the definition of the b-values used in the DWI acquisition. ADC calculation can only be performed using AT LEAST 2 b-values, as the ADC is the slope of signal loss between two different b-value acquisitions. Therefore, please do not state things like “…showed that ADC values obtained with a b factor of 300 s/mm²…” (Results, Parameters, last paragraph, but similar things are mentioned throughout the manuscript), as they are, if I understand it correctly, calculated from b-values 0 AND 300 s/mm². Similar things apply for the ADC values calculated from b factors 500 and 800, which are actually calculated from 0 AND 500, and 0 AND 800, respectively. In the materials and methods you state that “DWI was acquired using three different b factors”, but you forget to mention that b=0 s/mm² is (probably) also measured. It is unclear to me at the moment whether you used one acquisition with 4 b-values (0, 300, 500, and 800), or 3 different acquisitions (b-values 0 and 300, 0 and 500, 0 and 800, separately). If it is the second case, please specify explicitly if you used the same TE for each of the acquisitions.

Sorry we did not make this clear. Obviously one cannot calculate ADC without at least 2 b-values. We have not made this clear in the text. Four b-values with all the same TR/TE.

18/ The authors should specify why these b-value combinations were used. It is the general consensus that when calculating b-values only including b-values below 100-150 provides ADC values which are maximally weighted towards perfusion influences, whereas using only b-values above 200-300 nearly only contain diffusion and cellularity effects. As the authors (rightly) state that increased perfusion might conflict with the reduced diffusivity caused by higher cellularity, I cannot understand why they have not calculated ADC values from b-value sets that maximize these differences, fi. One ADC value from b-values 0 and 100, and one ADC value from b-values 300 and 800. In this way it could provide actual information on the different contributions. If I understand it correctly, none of the b-value sets in the current study ((0, 300), (0, 500), and (0, 800)) contain any of these influences maximally. It is true that (0, 300) will contain more perfusion effects, and (0, 800) will contain less, but not to such extent that you can separate the contributions. This should be examined and discussed in greater detail, as this is one of the things this study could provide that would add to the current literature knowledge.

We totally agree with you regarding this shortcoming of this study. When we started this study we understood that the lower b-values would suffer from perfusion effects and we tried to minimize it by looking at the higher values. B-values below 300 were already studied by Bozgeyik et al. We did not expect that the most sensitive value would be at 300. We are not able to go back and rescan this current dataset with lower b-values but are actively rescanning a new patient population with more b-values in the lower as well as higher b-value ranges.
19/ Discussion, page 12: “Zulkif [14] and Abdel [18] also found that a b-factor of
300s/mm2 was the most useful for differentiating benign or malignant lesions”.  
Actually, in paper [18] Abdel mentions that the ADC calculated from b-values 0, 
250, and 500 has a good discriminatory value, but they don’t mention that 300 (or 
even 250) was the best b-value to use. If possible, I would also prefer the use of 
the authors’ last names when referenced in the discussion.
We have fixed the author’s last name issue. Both Abdel and Bozgeyik who studied DWI 
at these lower b-values. It was Zulkif Bozgeyik’s reference that included b-value=300. 
We have not modified the text to reflect that.

20/ Discussion, page 13: “Although at higher b-values the sensitivity to perfusion
is reduced, our results show no difference in ADC values between benign and 
malignant groups with b-values of 500 or 800 s/mm2”. As mentioned above, ADC 
values calculated from b-values 0 and 500, or from 0 and 800, still contain a 
large contribution from the perfusion due to the inclusion of the 0 in the 
calculation. For that reason, you should examine the ADC values calculated from 
b-values 300 and 800, and from b-values 500 and 800 to see if these 
perfusion-insensitive measures can provide more information.  
This is an excellent idea but unfortunately this data was acquired at another hospital 
where we worked previously and we have no longer affiliated with that hospital nor do 
we have access to the raw data stored there.

21/ Table 2: the standard deviation of the max nodular diameter of the malignant 
lesions (0.41) does not exactly match the one given in Table 1 (0.46), please 
correct. Also, indicate the units of the ADC values in the Table. One thing that is 
striking in this table is that the ADC (b-300) in the malignant group is lower than 
the ADC (b-500), as this is theoretically impossible. This should be checked 
carefully, and the reason should be found/mentioned.  
We have corrected the typo.  
Thank you for pointing out that anomaly of the ADC value. We realize that theoretically, 
ADC values computed using the standard single compartment model: ADC= 
\[-\ln(S(b)/S(0))/b\] should go down with increasing b-values. We suspect that the complex 
composition of cellularity of the malignant lesions present a multi-compartment situation 
where the standard equation is less accurate as higher b-values would be more 
sensitized to smaller compartments or distances.

22/ The authors should attempt to provide a clearer view on the improvements 
that the current manuscript introduces with regards to the three main referenced 
similar papers (references 14, 15, and 16). As the manuscript is written now, you 
have nearly identical purposes and conclusions, have included less patients 
(albeit slightly more malignant lesions), and found similar or worse sensitivities 
and specificities. The main contribution you can make to the current knowledge is 
that your reference standard is better than the biopsy that was used in most of 
the other studies, and that you use different b-values. In order to maximize the 
information in this work, you should focus more on the improvement that those
differences (might) give, rather than echoing the conclusions drawn in the previous studies. I would suggest comparing FNAB results to your golden standard, and recalculating the ADC values from different b-value combinations to maximize separation of perfusion and diffusion influences. As suggested before we did not expect that the most sensitive b-value combination would be the one at (0,300). As you indicated we did emphasize the uniqueness of our dataset which used the absolute gold standard of the pathology results. We do not have FNAB results on this cohort of patients to make any comparison. The anomalous results at (0,300) is what prompted us to continue this study with more b-values around this b-value as well as higher so we can look at multi compartment models’ effects on ADC. It is not always possible to obtain patients’ data with both pathology readings and DWI, so this is a slow project.

Level of interest: An article of limited interest
Quality of written English: Needs some language corrections before being published
Statistical review: No, the manuscript does not need to be seen by a statistician.
Declaration of competing interests:
I declare that I have no competing interests
Reviewer's report 3
Title: Diagnostic value of Diffusion-Weighted MR imaging in thyroid disease: application in differentiating benign from malignant disease
Version: 2 Date: 7 December 2012
Reviewer: Ahmed Abdel Razek
Reviewer's report: the manuscript needs major corrections

Introduction:
• OK
• Material and methods:
  • Better defined the inclusion and exclusion criteria.
  Inclusion criteria: patients with thyroid nodules diagnosed by ultrasound (US) were prospectively included in the study. All patients had lesions larger than 1 cm in the greatest minimal transverse diameter. All enrolled patients have been scheduled to have thyrodecomy within two weeks. Exclusion criteria: patients with bad MR image quality and motion artifacts were excluded. Nodules less than 1 cm in size were not included the study.

• Results:
  • Add table for the ADC value of benign and malignant lesions at different B value with their sensitivity, specificity and AUC:
    We have added the table as Table 3 in the manuscript.

• Discussion:
  • There is bias from selection of B value less than 500 because it is associated with perfusion effect.
    We have acknowledged this in the text and also because the fact that at b=300 we had the most significant results similar to other reports, we hypothesized that may be a complex interplay of perfusion and diffusion effects provided a signature for DWI at 300. We are in the process of investigating this phenomenon with more b-values around 300 and also higher b-values to look at multi-compartment effects of malignant lesions. We currently do not have other b-values on this patient cohort.

• Ref:
  • Ref 12, 18 is the same
    We have corrected this now.

• Figures:
  • Add arrow for the lesion and ROI on one lesion
    We have now indicated that on the figs.

  • You can include images of ADC map beside DWI
    We have included the ADCs map in the figs.
• Figure 1 is cyst did you measure in the cyst or what is the ROI localization
  Figure 1 showed a thyroid adenoma in right lobe, non-contrast and contrast T1 images
  showed hemorrhage in right lobe, so we placed ROI on the right upper corner to avoid
  the hemorrhage areas.

• It is better to put solid lesion
  We understand that it better to use a uniform sample but due to limited number of
  patients with pathology results the choices are limited.

• Figure 2 shows lesion of the left side of thyroid gland and adjacent lesion. It is
  better to localize the
  This sentence was not complete in the pdf that was forwarded to us. But, we have now
  updated the figures.

  Level of interest: An article of importance in its field
  Quality of written English: Acceptable
  Statistical review: No, the manuscript does not need to be seen by a
  statistician.
  Declaration of competing interests:
  I declare that I have no competing interests