Reviewer's report

**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.

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?

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.

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.

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

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).

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.”

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

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?

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

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.

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

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.

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.

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

16/ References: Reference 12 and 18 are identical.

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.

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, i.e. 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.

19/ Discussion, page 12: “Zulkif [14] and Abdel [18] also found that a b-factor of 300 ms/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.

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 ms/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.

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.

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.

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