Does diffusion-weighted imaging (DWI) add diagnostic confidence in discriminating between malignant and benign solid focal lesions if included in contrast-enhanced magnetic resonance imaging (MRI) of the liver?

Poster No.: C-2022
Congress: ECR 2012
Type: Scientific Paper
Authors: R. Girometti, M. Del Pin, S. Pullini, L. Cereser, G. Como, M. Bazzocchi, C. Zuiani; Udine/IT
Keywords: Abdomen, MR-Diffusion/Perfusion, MR, Decision analysis, Liver
DOI: 10.1594/ecr2012/C-2022

Any information contained in this pdf file is automatically generated from digital material submitted to EPOS by third parties in the form of scientific presentations. References to any names, marks, products, or services of third parties or hypertext links to third-party sites or information are provided solely as a convenience to you and do not in any way constitute or imply ECR’s endorsement, sponsorship or recommendation of the third party, information, product or service. ECR is not responsible for the content of these pages and does not make any representations regarding the content or accuracy of material in this file.

As per copyright regulations, any unauthorised use of the material or parts thereof as well as commercial reproduction or multiple distribution by any traditional or electronically based reproduction/publication method is strictly prohibited.

You agree to defend, indemnify, and hold ECR harmless from and against any and all claims, damages, costs, and expenses, including attorneys' fees, arising from or related to your use of these pages.

Please note: Links to movies, ppt slideshows and any other multimedia files are not available in the pdf version of presentations.

www.myESR.org
Purpose

Background

- The added value of Diffusion-weighted Imaging (DWI) in characterizing focal liver lesions (FLLs) is under debate. Diagnostic accuracy for malignancy widely varies in different studies [1], depending on lesion histology or technical factors that account for a large variability in the Apparent Diffusion Coefficient (ADC) values and signal intensity [2-3]. Accordingly, the differentiation between malignant and solid benign FLLs (as hepatocellular adenoma (HA) or focal nodular hyperplasia (FNH)) is challenging [2].
- Less defined is the role of DWI in the diagnostic thinking process of radiologists. To our knowledge, no previous studies investigated whether DWI impacts on the confidence with which radiologists assess FLLs. On the other hand, diagnostic confidence has a crucial role on medical decision-making, because further investigations or therapeutic actions are often based on levels of confidence [4]. We wondered whether DWI has the potential to impact on the diagnostic confidence in assessing solid FLLs, particularly in the case of correct diagnosis.

Purpose

The purpose of this study was to investigate whether adding DWI to conventional Magnetic Resonance Imaging (MRI) of the liver increases radiologists diagnostic confidence in assessing FLLs.
Methods and Materials

Patients population

- Included were 43 subjects (21 male, 22 female; age range 21-79 years, mean age 57.7 years) referred to MRI (February 2009-June 2011) for first assessment of sonographically-detected FLLs. Patients showed hepatic cirrhosis in 19 cases and a non-cirrhotic liver in 24 cases. A total of 69 lesions were included in the analysis because of a maximum diameter ≥1.0 cm, in order to provide adequate lesion conspicuity on ADC maps in the image analysis.

- Excluded were patients: (i) with poor DWI image quality; (ii) in whom the standard of reference for final diagnosis was not available; (iii) showing typical MRI appearance of cysts and/or haemangiomas [5].

Standard of reference

- 21 lesions: histopathologic analysis after biopsy or surgery (Table 1 on page 6).
- 48 lesions: verification by 1-year follow-up CT and/or MRI of a consensus panel diagnosis obtained by two experienced readers (C.Z., M.B.) according to standardized criteria [5-6] (Table 1 on page 6). Lesions that remained stable in number and dimensions despite the absence of therapy were assumed to be benign (n=11), including 3 lesions in cirrhotic livers finally assessed as dysplastic nodules. Lesions were assumed to be malignant when presenting: (i) suspected appearance at baseline MRI followed by increase in number and dimensions and/or response to medical therapy or interventional procedures (n=8); (ii) a typical pattern for HCC according to the recently validated criteria of the American Association for the Study of Liver Diseases [6] (n=19). Regardless of specific appearance, FNHs and HAs were included in one single group of benign solid FLLs, due to potential difficulties in differential diagnosis by imaging alone [7]. Metastases were secondary to esophageal cancer in 5 cases, pancreatic cancer in 3 cases, breast cancer in 2 cases, and colon cancer in two remaining ones.

MRI protocol

- Examinations were performed on a 1.5T magnet (Magnetom Avanto, Siemens Medical Systems, Erlangen, Germany) using a surface phased-array body coil. The study protocol is resumed in Table 2 on page 6.
- DWI was performed before the dynamic study with a respiratory-triggered Single-Shot Echo-planar sequence (SS SE EPI) acquired on the axial plane, with b values of 50, 500 and 800 sec/mm² (Table 2 on page 6). In order to reduce artefacts, we used Fat saturation with SPAIR technique and a
parallel imaging algorithm (Generalized Autocalibrating Partially Parallel Acquisition; GRAPPA) with an acceleration factor of 2.

**Image analysis**

- Performed on a dedicated console (Leonardo Syngo, Siemens Medical Systems, Erlangen, Germany) by two readers (G.C., L.C.) blinded to clinical information with 12 and 5 years of experience in abdominal radiology, respectively (R1 and R2). Each MRI examination was independently presented in two different sets: (i) set A included the conventional MRI protocol without DWI; (ii) set B included the conventional protocol with DWI images (at corresponding b-values) and the ADC map. Lesions and liver ADCs were already calculated by a third radiologist, according to the method resumed in Fig. 1 on page 7. Sets presentation was separated by an interval of time of four weeks in order to avoid recall bias. Multiple lesions in a same patient were randomly presented as separate images for both sets.
- Readers were asked to discriminate between malignant and benign FLLs, referring to well known criteria for T1, T2 and post-contrast imaging [5-6]. DWI features for restricted diffusivity rising the suspicion for malignancy included: lesion hypointensity on the ADC map (with or without hyperintensity persisting from b = 50 to 800 sec/mm$^2$ on DWI images) [8], and/or a lesion ADC one or more standard deviation below the ADC of the surrounding liver parenchyma. In order to reflect the complexity of the clinical scenario, readers were left free on how to combine information from DWI/not-DWI data and visual analysis/ADC value in providing set B diagnoses.

**Analysis of confidence**

- **Confidence score**: for sets A and B, readers were asked to express how they were subjectively confident in each single diagnosis (malignancy or benignancy).
- **Usefulness score**: for the set B only, readers were asked to express the confidence in the contribution of DWI to overall diagnosis, i.e. the impression that DWI was useful to make the diagnosis.
- Both scores were expressed on a 1 to 5 scale, with 1 standing for "none" (confidence, impact or usefulness quantifiable in 0-20%), 2 for "low" (21-40%), 3 for "intermediate" (41-60%), 4 for "high" (61-80%), and 5 for very high (81-100%).

**Statistical analysis**

- Data analysis was performed on a per-lesion basis.
- Intra- and inter-reader differences in confidence and usefulness scores were assessed with the Wilcoxon ranked sign test and Mann-Whitney u test, respectively.
• Based on the comparison with the standards of reference, we calculated the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and accuracy for malignancy of both readers. By assuming that the diagnostic confidence was "low" for scores 1-3 and "high" for scores 4-5, we also estimated the proportion of correct diagnoses for each confidence level and for each image set. Exact binomial 95% C.I.s were calculated for all proportions.
• Alfa value = 0.05 was considered statistically significant.
Table 1: Characteristics of focal liver lesions in the study. Consensus diagnosis was provided by two expert radiologists who reviewed patients clinical data and imaging examinations performed over a 1 year follow-up (see text). HCC = hepatocellular carcinoma; FNH = focal nodular hyperplasia; HA = hepatocellular adenoma.

© Institute of Diagnostic Radiology, University of Udine - Udine/IT

<table>
<thead>
<tr>
<th>Final diagnosis</th>
<th>Number of lesions</th>
<th>Mean /maximum diameter (cm)</th>
<th>Histology</th>
<th>Consensus diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metastasis</td>
<td>20</td>
<td>1.2/3.7</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>HCC</td>
<td>31</td>
<td>1.0/7.5</td>
<td>7</td>
<td>24</td>
</tr>
<tr>
<td>Dysplastic nodule</td>
<td>3</td>
<td>1.2-1.5</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>FNH/HA</td>
<td>14</td>
<td>1.3/10.4</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Atypical haemangioma</td>
<td>1</td>
<td>1.4</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Sequence</td>
<td>Plane</td>
<td>TR/TE (msec)</td>
<td>Field of View (mm)</td>
<td>Acquisition Matrix</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-------</td>
<td>-------------------------</td>
<td>--------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>GE T1w in-/out of-phase</td>
<td>axial</td>
<td>100/2.35 and 5.04</td>
<td>300x400</td>
<td>512x384</td>
</tr>
<tr>
<td>STIR T2w</td>
<td>axial</td>
<td>3770/134</td>
<td>300x400</td>
<td>512x384</td>
</tr>
<tr>
<td>SPAIR T2w</td>
<td>axial</td>
<td>4510/65</td>
<td>285x380</td>
<td>320x240</td>
</tr>
<tr>
<td>DWI (SS SE EPI)</td>
<td>axial</td>
<td>1900/69</td>
<td>380x285</td>
<td>115x192</td>
</tr>
<tr>
<td>VIBE T1w</td>
<td>axial</td>
<td>4.2/1.4</td>
<td>337x400</td>
<td>256x216</td>
</tr>
</tbody>
</table>

* Multi-breath-hold
** Respiratory-triggered
*** Single breath-hold

Table 2: Study protocol, including the following sequences: Gradient Echo (GE) T1-weighted, Short Time Inversion Recovery (STIR) or respiratory-triggered Spectral Adiabatic Inversion Recovery (SPAIR) T2-weighted, and a volumetric, fat-saturated Volume Interpolated Breath Hold Examination (VIBE) T1-weighted. VIBE was used for the dynamic study, before and after intravenous injection of 0.1 mL/Kg of gadobenate dimeglumine (MultiHance, Bracco, Milan, Italy). The sequence was repeated on the axial plane after 30, 70 and 300 sec after contrast administration. Images were acquired also in the hepatospecific phase at about 1 hour after contrast administration.

© Institute of Diagnostic Radiology, University of Udine - Udine/IT
Fig. 1: Measurements of ADC were performed by positioning as large as possible circular regions of interest (ROIs) on $b=50$ sec/mm$^2$ images. One single as large as possible ROI was placed to cover entirely a lesion in the slice showing the largest diameter, by excluding from measurements the surrounding parenchyma or artifacts. In the case of inhomogeneous lesions, 1-3 as large as possible ROIs (depending on lesion dimensions) were positioned avoiding areas of necrosis and/or haemorrhage and/or including solid areas with different signal intensity. The ADC was calculated by transporting the ROIs on the ADC map with a copy and paste operation, and averaging the measurements in case of multiple ROIs. In order to compare the ADCs of FLLs with that of the surrounding parenchyma, 3 additional ROIs were positioned around a lesion (minimum distance of 1 cm) in the same $b=50$ sec/mm$^2$ slice. As in the case of FLLs, parenchymal ADC was calculated by averaging measurements obtained by transporting the ROIs on the ADC map.

© Institute of Diagnostic Radiology, University of Udine - Udine/IT
Results

Analysis of confidence

- By analyzing set A images (conventional MRI alone), both readers were shown to be highly confident in discriminating between malignant and benign FLLs, with overlapping confidence scores of 4.01±1.06 for R1 and 4.16±0.97 for R2 (p>0.05). As shown in Table 3 on page 11, adding DWI to the conventional MRI protocol did not significantly modify the confidence score of R2 (p>0.05), while determined a significant decrease in the diagnostic confidence of R1 (p=0.0184). R1 (the more experienced reader) was significantly less confident in his set B diagnoses as compared to R2 (the less experienced one) (p=0.0236).

  When analyzing set B images, R1 attributed a higher usefulness score (3.04±0.96) to DWI as compared to R2 (2.71±1.20), although the difference was not statistically significant (p>0.05) (Table 3 on page 11).

Diagnostic performance

- The diagnostic performance of the two readers is resumed in Table 4 on page 11. The proportion of correct diagnoses with low or high diagnostic confidence for the sets A and B is detailed in Table 5 on page 12. Overall, from set A to set B we found an increase in the proportion of less confident correct diagnoses (+11.6 and +10.2% for R1 and R2 respectively), and a decrease in the proportion of highly confident correct diagnoses (-10.1 and -7.2% for R1 and R2, respectively).

- Readers had same false-negative (FN) lesions (Fig. 2 on page 13) without and with the use of DWI (Table 6 on page 14). R1 was less sensitive than R2 (88.2 vs. 96.1%, respectively) because of 6 FNs, corresponding to 5 HCCs without typical contrast-enhancement pattern (no wash-in and wash-out dynamic) and 1 metastasis with incomplete contrast-enhancement. The average confidence score for these lesions was higher reading the set B rather than set A images (3.0-4.5 vs. 2.0-3.5, respectively), while DWI was considered of intermediate usefulness in making the diagnosis (score 3.0-3.5). R2 had two FN HCCs in common with reader 1 (Tab. 5). However, no difference was found in the average confidence score (4.0) with or without DWI. R2 attributed minimal usefulness to DWI in making FN diagnoses (average score of 1.5).

- By analyzing the set B, readers avoided three false-positive (FP) lesions, changing set A diagnosis from: (i) metastasis to atypical haemangioma in one case for reader 1 (confidence score of 4 in both sets; usefulness score of 5) (Fig. 3 on page 15); (ii) metastasis to adenoma in 2 cases for reader 2 (average confidence score from 2 to 3; usefulness score of 4). Accordingly, DWI added specificity, PPV and NPV to conventional MRI, although the overlap between the 95% C.I.s suggests no statistically
significant difference. For R2, 5 FP remained unchanged from set A to set B. As evident from Table 6 on page 14, the average confidence scores for set B attributed by R2 to these cases were similar or lower to those of the set A, whereas the usefulness score was lower (1.0-1.5).
Table 3: Confidence scores attributed by readers to the diagnosis (malignancy or benignancy) of focal liver lesions obtained with set A (conventional MRI) or set B images (conventional MRI + Diffusion-weighted Imaging). Usefulness score is reported also.

© Institute of Diagnostic Radiology, University of Udine - Udine/IT

<table>
<thead>
<tr>
<th></th>
<th>Set A</th>
<th>Set B</th>
<th>p</th>
<th>Usefulness score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>READER 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>range mean ± SD</td>
<td>2-5</td>
<td>2-5</td>
<td>0.0184</td>
<td>2-5</td>
</tr>
<tr>
<td></td>
<td>4.01±1.06</td>
<td>3.72±0.87</td>
<td></td>
<td>3.04±0.96</td>
</tr>
<tr>
<td><strong>READER 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>range mean ± SD</td>
<td>2-5</td>
<td>2-5</td>
<td>N.S.</td>
<td>2-5</td>
</tr>
<tr>
<td></td>
<td>4.16±0.97</td>
<td>4.08±0.98</td>
<td></td>
<td>2.71±1.20</td>
</tr>
<tr>
<td><strong>p</strong></td>
<td>N.S.</td>
<td>0.0236</td>
<td></td>
<td>N.S.</td>
</tr>
</tbody>
</table>
Table 4: Diagnostic performance of the readers in assessing focal liver lesions malignancy with set A or B images. 95% C.I. are reported in brackets. PPV = positive predictive value; NPV = negative predictive value.

© Institute of Diagnostic Radiology, University of Udine - Udine/IT
Table 5: Changes in the proportion of readers' correct diagnoses of malignancy between the sets A and B. Proportions are calculated on the total of lesions (n=69).

© Institute of Diagnostic Radiology, University of Udine - Udine/IT
**Fig. 2:** False-negative case for malignancy of R1 in a seventy-six years-old male patient with viral chronic hepatitis and hepatocarcinoma (readers were blinded to clinical information). Focal lesion in the left liver lobe showed signal intensity suggesting solid, hypercellular nature, as evident from GE out of-phase T1-weighted (a) and SPAIR T2-weighted (b) images, respectively. The VIBE T1-weighted sequence after i.v. gadolinium injection showed contrast wash-in in the arterial phase (c) without wash-out during portal (not shown) and equilibrium (d) phases. Hypointensity on hepatospecific phase was observed (not shown). Hyperintensity on b = 800 sec/mm² image (e) and ADC map (f) suggested preserved diffusivity and co-existence of the T2-shine-through effect.
Table 6: False-negative and false-positive cases for malignancy obtained both with set A and set B images. "Hypo", "hyper" and "iso" are abbreviations referring to lesion hypointensity, hyperintensity or isointensity as compared to the surrounding parenchyma, respectively. In case of multiple lesions, average scores are reported. MRI = Magnetic Resonance Imaging; DWI = Diffusion-Weighted Imaging; ADC = Apparent Diffusion Coefficient.

© Institute of Diagnostic Radiology, University of Udine - Udine/IT
**Fig. 3:** False-positive case for malignancy in a sixty-six years-old female patient operated for endometrial carcinoma. One focal lesion (arrow) was visible in the right hepatic lobe, appearing as hypointense on the GE in-phase T1-weighted image (a) and slightly hyperintense on Short Time Inversion Recovery (STIR) T2-weighted image (b). No contrast uptake was found, except for a thin peripheral rim enhancement, as shown in the VIBE T1-weighted image acquired in the equilibrium phase. R2 assessed the lesion as a metastasis. However, DWI showed a large decrease in signal hyperintensity moving from $b = 50$ (d) to $b = 800$ sec/mm$^2$ (e) images, and hyperintensity on the Apparent Diffusion Coefficient map (f), suggesting preserved diffusivity and an underlying T2-shine-through effect (arrow).

© Institute of Diagnostic Radiology, University of Udine - Udine/IT
Conclusion

Limitations

• We investigated a relatively small sample of FLLs because of the exclusion of cysts and haemangiomas from analysis. Our choice was based on the assumption that the diagnosis of these lesions is easy with conventional MRI, while the main contribution of DWI to the clinical practice should more properly consist in helping to differentiate between solid benign vs. malignant FLLs [2].

• No definite ADC threshold for malignancy was provided to readers, leading to: (i) more uncertain interpretation of part of the DWI information; (ii) higher weight of visual analysis in defining the diagnosis as compared to ADC quantification. However, because of the variability in ADC values due to equipment- and study protocol-related factors [9], our study reflects the clinical practice, in which no reliable threshold is generalizable, and lesions ADC should be compared to that of the surrounding parenchyma to provide as more as possible reliable information in combination with visual analysis results [10].

• Readers were blinded to clinical information, which is often crucial in assessing FLLs. This might have led to an underestimation of readers' diagnostic performance, especially in terms of sensitivity for R1 and specificity for R2. However, blinding was indispensable to avoid that clinical information might act as a confounder in data interpretation.

Main conclusions

• DWI didn't change or reduced global radiologists' confidence in differentiating between solid benign and malignant FLLs as compared to conventional MRI (Table 3 on page 19). However, the use of DWI was associated with a decrease in the confidence with which radiologists made correct diagnoses (Table 5 on page 19). Thus, T1- and T2-weighted images, together with contrast dynamic study (set A), suffice to provide maximal diagnostic confidence.

• Our results are probably related to the lack of a standard in interpreting and combining information from visual analysis and ADC quantification, leading to an increase (for the more experienced reader) or a "steady state" (for the less experienced reader) in uncertainty when matching DWI and conventional MRI data within a unified diagnosis (Table 3 on page 19, Table 5 on page 19). We assumed that the more experienced reader was able to combine information from conventional MRI and DWI more easily than R2, leading to the subjective impression of higher usefulness of DWI. This impression paradoxically corresponded to a lower diagnostic confidence when reading set B images, suggesting that the additional information from DWI can act as a confounding factor if a confident, expert
radiologist interprets well-established [5-6] MRI features. In the case of the less experience reader, who is assumed to be more globally uncertain [11], DWI was probably less useful both in confirming or rejecting hypotheses based on conventional MRI findings, leading to similar diagnostic confidence with or without DWI.

- Our results suggest that when conventional MRI does not suffice to provide a diagnosis, DWI is of no help in addressing the correct one (Table 6 on page 20). In the case of the more experienced reader, DWI increased the confidence in false-negative and false-positive cases with atypical appearance on conventional MRI, suggesting that advantages from DWI might be impaired by the potential to significantly contribute to incorrect diagnoses. Giving the tendency of the more expert reader to find DWI more useful than R2 in making the diagnosis, it is not surprising that he used additional information to increase his confidence in incorrect diagnoses. On the contrary, R2 didn’t show a definite trend of confidence in those false-positive cases that remained unchanged between the sets A and B. Regardless of lesions appearance both on conventional MRI and DWI, the reader attributed a minimal usefulness score to DWI, even when visual analysis and ADC values suggested benignancy in 2 dysplastic nodules and 1 atypical haemangioma. Thus, he probably based his diagnostic decision on conventional MRI findings only, in accordance with the above hypotheses.
Table 3: Confidence scores attributed by readers to the diagnosis (malignancy or benignancy) of focal liver lesions obtained with set A (conventional MRI) or set B images (conventional MRI + Diffusion-weighted Imaging). Usefulness score is reported also.

© Institute of Diagnostic Radiology, University of Udine - Udine/IT
Table 5: Changes in the proportion of readers' correct diagnoses of malignancy between the sets A and B. Proportions are calculated on the total of lesions (n=69).

© Institute of Diagnostic Radiology, University of Udine - Udine/IT
Table 6: False-negative and false-positive cases for malignancy obtained both with set A set B images. "Hypo", "hyper" and "iso" are abbreviations referring to lesion hypointensity, hyperintensity or isointensity as compared to the surrounding parenchyma, respectively. In case of multiple lesions, average scores are reported. MRI = Magnetic Resonance Imaging; DWI = Diffusion-Weighted Imaging; ADC = Apparent Diffusion Coefficient.

© Institute of Diagnostic Radiology, University of Udine - Udine/IT
References


Personal Information

Thank you for the interest in this poster.
Do not hesitate to contact me for any question.

rgirometti@sirm.org