Comparison of qualitative and quantitative approach to prostate MR spectroscopy in peripheral zone cancer detection

Stijn Klijn, Pieter J. De Visschere, Gert O. De Meerleer, Geert M. Villeirs

Division of Genitourinary Radiology, Ghent University Hospital, De Pintelaan 185, 9000 Gent, Belgium
Division of Radiotherapy, Ghent University Hospital, De Pintelaan 185, 9000 Gent, Belgium
Radboud University Nijmegen Medical Centre, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands

ARTICLE INFO
Article history:
Received 29 September 2010
Accepted 2 December 2010

Keywords:
Magnetic resonance imaging
Magnetic resonance spectroscopy
Prostatic neoplasms
Diagnostic techniques and procedures
Image interpretation

ABSTRACT
Objective: To compare the diagnostic performance of a qualitative (pattern recognition) and a quantitative (numerical assessment) approach to magnetic resonance spectroscopy (MRS) in the diagnosis of peripheral zone prostate cancer.

Methods: 185 patients (131 with histopathologically proven cancer, 54 normal/benign after at least 12 months follow-up) were prospectively evaluated with qualitative MRS using a 4-point scale between 3/2004 and 1/2008, and retrospectively reassessed using a prototype quantitative postprocessing software in April 2008. Based on pathology and follow-up data, diagnostic performance parameters were calculated.

Results: The qualitative and quantitative approaches were concordant in 78.9% (146/185) of cases. The difference between the areas under the ROC curve (0.791 versus 0.772, respectively) was not statistically significant. The sensitivity, specificity and accuracy were 55.7%, 94.4% and 67.0% for the qualitative approach, and 55.0%, 83.3% and 63.2% for the quantitative approach. The sensitivity for high grade tumours (Gleason 4+3 or higher) was 85.2% (23/27) for both approaches. All cancers missed on either one approach separately (31/31) and 91% of cancers missed on both approaches together (23/27) were of lower grade (Gleason 3+4 or lower).

Conclusions: Qualitative and quantitative approaches to MRS yield similar diagnostic results. Discordances in tumour detection only occurred in lower grade cancers.

© 2010 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Magnetic resonance spectroscopy (MRS) provides metabolic information from voxels throughout the prostate. When healthy citrate-producing cells transform into malignant citrate-oxidizing cells, they lose the ability to accumulate and secrete very high levels of citrate into the ductal lumen [1]. Concurrently, the concentration of choline (inolved in the cellular membrane synthesis and degradation) increases due to altered choline transport, increased membrane catabolic activity, and/or increased choline kinase activity [2,3]. MRS can thus detect prostate cancer in areas with decreased citrate and increased choline concentrations. It has proven to be helpful in localizing prostate cancer, increasing specificity, and assessing tumour aggressiveness [4–6].

In a recent meta-analysis, the sensitivity and specificity of combined magnetic resonance imaging (MRI) and magnetic resonance spectroscopy on a patient level has been found to be 82% and 88%, respectively [7]. Most of the included reports were based on mathematical curve fitting and quantification of metabolite concentrations, and the calculation of metabolite ratios using computer software [4,8–18]. This technique, however, involves time and expertise to post-process and quality-assure the fitting results. It has a steep learning curve, is generally perceived to be quite tedious and time-consuming, and has therefore not yet received broad clinical acceptance by radiologists.

Alternatively, a qualitative approach to the interpretation of MR spectra can be performed, based on pattern recognition through direct visual estimation and correlation of choline and citrate peak heights [19,20]. This method seems faster and easier to apply in the majority of cases, and was reported to yield competitive diagnostic performance as compared to reported quantitative analyses [19]. It might therefore help decreasing the reluctance of many radiologists to add MR spectroscopy to the standard imaging protocol. To the best of our knowledge, however, no direct comparison between qualitative and quantitative spectroscopic evaluation in the same patient group has been reported so far. The goal of this study was to compare the performance of both approaches in the detection of peripheral zone prostate cancer in a large group of patients and to identify caveats and sources of error influencing their diagnostic performance.
2. Materials and methods

2.1. Patients

We identified 185 patients (mean age 63.2 years, range 44–80 years; mean PSA 12.7 ng/ml, median 8.3 ng/ml, range 0.69–133 ng/ml) who underwent combined endorectal coil MRI and MRS with qualitative evaluation at our institution, and of whom imaging files were digitally available for retrospective quantitative review. They either were referred for diagnosis on the basis of elevated PSA (no previous biopsy, \( n = 113 \)) or for pre-radiotherapy work-up (biopsy-proven prostate cancer, \( n = 72 \)).

The study was approved by our hospital’s Ethics Committee (Project EC UZG 2007/459).

2.2. MRS technique

All patients were examined on a 1.5T scanner (Magnetom Symphony, Siemens Medical Systems, Erlangen, Germany). An endorectal coil (MRInervu, Medrad, Pittsburgh, USA) inflated with 60 ml of air was combined with a pelvic phased-array coil, tightly wrapped around the pelvis. Before coil insertion, a laxative enema or suppository was administered to empty the rectum, and an intravenous spasmolytic drug was injected to avoid bowel movements and facilitate coil insertion. The acquisition protocol consisted of morphologic fast-T2-weighted sequences (4-mm thickness without interslice gap, transverse, coronal and sagittal orientation, TR/TE 4000–4400/139 ms, 180° flip angle, 14 cm × 18 cm field of view, matrix size 180 × 512, 2 acquisitions, acquisition time 3–4 min) and 3-dimensional magnetic resonance spectroscopic imaging. A point-resolved spectroscopic sequence (PRESS) was used with TR/TE 650/120 ms, bandwidth 1250 Hz, vector size 512 samples [21]. The field of view, matrix size and number of weighted acquisitions was adapted to the prostate size to achieve a nominal voxel size of 0.25 cc or smaller in a total measurement time of 10 min. Unwanted water and lipid signals in the region of interest were reduced by applying spectral selective suppression of water and lipid signals, together with 7 selective outer-voxel saturation bands to reduce contamination from surrounding structures, especially periprostatic fat and rectal air. They were positioned close to the prostate margins on transverse (1 posterior, 2 lateral and 2 anterior) and sagittal (superior and inferior) images to conform the prostatic shape as closely as possible. Shimming in the region of interest was performed automatically, followed by manual fine-tuning up to a line width of the water resonance peak of 30Hz or lower.

All data were prospectively assessed using a qualitative approach between March 2004 and January 2008. In April 2008, the same data were retrospectively assessed using a quantitative approach.

2.3. Qualitative approach

The raw spectral data were apodized using a Hanning filter (center 0 ms, width 512 ms) and Fourier transformed. Zero-filling was performed as needed to adapt the apparent spectral resolution to the corresponding MR imaging data, residual water was removed, and baseline correction (calculation range 2.0–4.0 ppm, excluded ranges 2.40–2.80 ppm and 2.95–3.30 ppm, polynomial order 6) and automatic phase correction were performed, with manual fine-tuning if needed.

Spectral maps, consisting of spectroscopic signals (display range 1.3–4.0 ppm) overlaid on the corresponding transverse T2-weighted images (Fig. 1), were obtained using Syngo software (Siemens Medical Systems, Erlangen, Germany). Citrate peaks were identified as doublets or single peaks (according to spectral quality) around 2.6 ppm, whereas choline and creatine peaks were recognized as single peaks around 3.2 and 3.0 ppm, respectively. With lower spectral quality or in case choline or creatine peaks were very high, both peaks were regarded as one peak [22]. Voxels that were of insufficient spectral quality, that contained glycerophosphocholine, or that showed partial volume effects or overlapping fat suppression bands, were ignored. Good spectral quality was defined as the ability to reliably distinguish citrate and total choline/creatinine peaks from background noise, in the absence of substantial lipid or water contamination [23]. Glycerophosphocholine is a normal constituent of fluid within the seminal vesicles and ejaculatory ducts, and is virtually indistinguishable from phosphocholine (the most important marker for prostate cancer) [24]. This may result in inappropriate detection of elevated choline peaks at the prostatic base, due to spectroscopic point spread function and partial volume averaging with the seminal vesicles, and along the trajectory of the ejaculatory ducts. Furthermore, peripheral voxels containing less than 50% prostatic tissue or voxels covered by fat suppression bands for more than 50% of their area were disregarded.

All qualitative evaluations were prospectively performed, by a single radiologist having more than 10 years of experience in prostate imaging. The spectroscopic data were scored through a 4-point scale [5] (Table 1): 1 = normal (citrate peak at least two times higher than choline/creatine peak height); 2 = probably benign (citrate peak smaller than two times or equal to choline/creatine peak); 3 = probably malignant (choline/creatine peak smaller than two times but higher than citrate peak); 4 = definitely malignant (choline/creatine peak at least two times higher than citrate peak height). Allocation to a higher diagnostic scale was only allowed when at least three adjacent spectroscopic voxels showed a similar spectroscopic profile (e.g. one isolated scale 4 voxel within an area of scale 2 voxels remained scale 2).

The examination time of the qualitative approach was measured in 10 randomly selected patients during daily clinical routine.

2.4. Quantitative approach

All intraprostatic spectroscopic voxels not covered by fat suppression bands were selected and postprocessed using a prototype software (Metabolite Report, Siemens Medical Systems, Erlangen,
Table 1
Four-point scale spectroscopic scoring system.

<table>
<thead>
<tr>
<th>4-Point scale</th>
<th>MRSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Normal</td>
<td>Choline peak at least two times higher than creatine peak</td>
</tr>
<tr>
<td>2 Probably benign</td>
<td>Choline peak smaller than two times or equal to creatine peak</td>
</tr>
<tr>
<td>3 Probably malignant</td>
<td>Choline/creatinine peak smaller than two times but higher than citrate peak</td>
</tr>
<tr>
<td>4 Definitely malignant</td>
<td>Choline/creatinine peak at least two times higher than citrate peak (if present)</td>
</tr>
</tbody>
</table>

Adapted with permission from reference [19]. Copyright Elsevier 2009.

Germany). The spectroscopic data were overlaid on the corresponding transverse T2-weighted images, showing the measured peaks (white), the calculated baseline (blue) and curve-fitted metabolite signals (red) (Fig. 2a). Using the areas under curve of the latter, choline plus creatine to citrate ratios (CC/C) of all intraprostatic spectroscopic voxels were calculated and displayed on a parametric color map equally overlaid on morphologic images (Fig. 2b). Voxel with an automatically calculated signal-to-noise ratio (SNR) less than 5 were not included and marked on the parametric image. Since prostate cancer correlates with high CC/C-ratios [6,22], quantitative spectroscopic analysis aims at detecting the highest CC/C-ratios. Analogous to the qualitative approach, a high CC/C-ratio was only considered diagnostically relevant when at least three adjacent spectroscopic voxels showed a high ratio. All numerical data were therefore transferred off-line and further evaluated with an in-house developed algorithm using Excel software (Microsoft, Seattle, USA). A list of highest CC/C-ratios in three adjacent voxels (triplets) was automatically created. To ensure location of the voxels within the peripheral zone and to distinguish diagnostic from inadequate voxels, all triplets were evaluated by a radiologist starting with the triplets with highest CC/C-ratios. Reasons for insufficient quality were low signal to noise ratio, inadequate fat suppression, inadequate water suppression, phase shifts and spectral degradation due to patient or coil movements, interference by metallic prostheses, or inadequate shimming (Table 2). Such conditions may result in a spectrum that cannot be adequately fitted: fitting of background noise peaks as small citrate or choline peaks due low SNR, incorrect fitting of citrate or choline peaks due to inadequate fat or water saturation, respectively, and inappropriate truncation of a broadened metabolite peak into choline and creatine components, or creatine and citrate components. If all three peripheral zone voxels were of sufficient quality and not influenced by partial volume effects from the seminal vesicles or ejaculatory ducts, the lowest CC/C-ratio of these three adjacent voxels was accepted as diagnostically most

Fig. 2. (a) Metabolite report spectroscopic map showing metabolite peaks overlaid on corresponding transverse T2-weighted MR image for quantitative analysis. Color codes: measured peaks in white, mathematical curve fits in red and baseline in blue. (b) Ratio map showing calculated ratios in every voxel. Background parametric color map shows lower ratios in blue and higher ratios in red. Red crosses indicate voxels with signal-to-noise ratio less than 5.
representative. If not, the next highest triplet of CC/C-ratios was assessed.

The examination time of the quantitative approach was measured in the last 10 patients of the retrospective quantitative study.

2.5. Pathology results and follow-up

Prostate cancer was histopathologically confirmed and graded after transrectal ultrasound-guided biopsy (n = 106) or radical prostatectomy (n = 25). The absence of prostate cancer (n = 54) was confirmed when no cancer was detected pathologically after at least 12 months and/or when serum PSA decreased or had a doubling time of at least 3 years [25].

2.6. Statistical analysis

On the basis of pathology and follow-up data, the diagnostic performance of the qualitative and quantitative approaches was assessed using the highest scale or ratio, respectively. Receiver operating characteristics (ROC) curves were plotted using SPSS software (SPSS for Windows, Rel. 15.0.0, SPSS Inc., Chicago, USA) and tested for significance of differences [26]. A p-value < 0.05 was considered to show statistical significance.

Assuming diagnostic scales 3 and 4 as indicative of malignancy, descriptive statistics (sensitivity, specificity, and accuracy) were calculated for the qualitative analysis. Using a CC/C cut-off that yields equal sensitivity for the qualitative and quantitative approach, the corresponding specificity and accuracy of the quantitative approach were calculated.

3. Results

The qualitative and quantitative evaluation yielded concordant results in 78.9% (146/185) (Table 3). These discordant results were correct in 101 cases (57 detected cancers and 44 true negative results) and incorrect in 45 cases (43 missed cancers and 2 false positive results). Ninety-one percent (39/43) of the missed cancers were lower grade (Gleason 3+4 or lower).

In case of discordance (n = 39), 15 cancers were missed by the qualitative approach and 16 by the quantitative approach (mean CC/C of missed cancers 0.74; range 0.50–0.93). All of them were lower grade cancers (Gleason 3+4 or lower). In addition, 87.5% (7/8) of false positive results were attributable to the quantitative approach (mean CC/C of false positive results 1.38, range 0.95–2.02).

The difference between the areas under the ROC curve for the qualitative (0.791) and quantitative approach (0.772) was not statistically significant (p = 0.32) (Fig. 3). The sensitivity, specificity and accuracy of the qualitative approach were 55.7%, 94.4% and 67.0%, respectively. For an equal sensitivity (55.0%), the specificity and accuracy of the quantitative approach were 83.3% and 63.2%, respectively, using a CC/C-ratio cut-off of 0.94. The sensitivity for high grade tumours was 85.2% (23/27) for both the qualitative and the quantitative approach. The mean CC/C-ratio for lower grade tumours was 2.87 (± 0.78 SD) and for high grade tumours 7.66 (± 0.94 SD).

The total evaluation time of the qualitative approach in normal clinical routine (generation of spectral maps using a fixed postprocessing protocol, quality assurance and interpretation of resulting data) was less than 5 min in all patients, and 2 min on average. The quantitative approach consisted of automatic generation of curve-fitted spectral maps and data files (3 min, runs in background), tracking of 3 adjacent voxels with highest metabolite ratios (1–2 min, runs in background), approval and interpretation of these ratios based on spectral quality and location in the peripheral
zone through correlation with the corresponding spectra, yielding a total evaluation time of 2–10 min per patient, and 4 min on average.

4. Discussion

A three-dimensional magnetic resonance spectroscopic examination of the prostate yields a huge amount of data, frequently in the range of over 1000 voxels per patient. Automated processing of these data by a computer therefore seems straightforward. For example, the likelihood of malignancy can be estimated by constructing confidence intervals above mean metabolite ratios in normal prostate tissue [22,23,27]. Color-coded parametric image overlays can be generated providing an intuitive analysis of metabolite concentrations and ratios throughout the prostate, enabling easy visualization of both tumour and normal areas. Quantification of spectroscopic data enables comparison of data on an intra- and interpatient basis, e.g. in the follow-up of prostate cancer during treatment [28]. But above all, computers are fast, never get tired, and make no mistakes, in contrast to humans who may overlook or misinterpret certain areas and who are subject to both intra- and interobserver variations.

On the other hand, computers do not directly assess the presence or absence of a tumour, but rather perform a predefined mathematical algorithm yielding a set of continuous data. Actually, they compare a given postprocessed metabolite peak in the frequency domain with prior knowledge files of how a corresponding standard metabolite peak looks like, and accordingly construct a mathematical curve that fits this given peak as closely as possible. On the basis of this fitted curve, measurements such as peak height, peak width, and area under curve can then be performed.

Accurate curve fitting is crucial to successful quantitative analysis of spectroscopic data, but highly relies on the quality of the postprocessed spectrum. In our quantitative study, a signal-to-noise cut-off of 5 was used to automatically exclude spectra of grossly unacceptable quality and the remaining voxels were deemed amenable to reliable curve-fitting. However, this was not invariably successful due to a number of errors that may result in erroneous quantification of resonance peaks (Table 2). Although the human eye can recognize and discard such errors, a computer can only follow the binary logic of a signal-to-noise cut-off and assign equal quality to all voxels above that cut-off, regardless of their SNR. Obviously, most errors occur in voxels in the lower SNR range, due to insufficient signal reception or low metabolic activity. The SNR decreases with increasing distance from the endorectal coil, e.g. at the anterolateral periphery of the prostate and just above or below the reception area of the endorectal coil; or in case of inappropriate coil positioning (too high, too low or rotated). Although many errors could be prevented by simply increasing the SNR cut-off (e.g. to 10 instead of 5), this would also lead to a considerable increase of discarded voxels.

Another reason for erroneous quantitative interpretation is the presence of glycercophosphocholine, a normal constituent of fluid within the seminal vesicles and ejaculatory ducts, resonating at a slightly higher frequency than phosphocholine (the most important marker for prostate cancer) [24]. Automatic curve fitting of total choline (between 3.10 and 3.30 ppm) does not differentiate glycercophosphocholine from phosphocholine. As a result, inappropriately high metabolite ratios can be found at the prostatic base (due to partial volume averaging with the seminal vesicles) and along the trajectory of the ejaculatory ducts.

Occasionally, we observed inappropriately fitted metabolite peaks even though the spectrum appeared to be of good quality on visual review. This was an inherent, although rare, problem of automatic curve fitting in this prototype postprocessing software, that was further fine-tuned in its later versions. Still other factors may jeopardize accurate quantitative spectroscopic evaluation, including inadequate phase and/or baseline-corrections, suboptimal metabolite modelling, quantification errors due to a different response to varying flip angles in the presence of B1 inhomogeneity and low metabolic activity. These factors, however, are expected to equally affect qualitative and quantitative evaluation and are therefore of minor importance for the present study.

Although quantitative techniques can greatly facilitate spectral assessment by automatically generating spectral peak information, ratios and color-coded parametric image overlays, they are not yet a “push on the button” technique, because voxel selection and quality assurance by an experienced radiologist remain needed. This might be obviated by future developments such as automatic detection and exclusion of water or lipid contaminated voxels, the search for an optimal SNR for exclusion of voxels of unacceptable quality, differentiation of phosphocholine and glycercophosphocholine containing voxels (and exclusion of the latter), improved metabolite peak modelling, automatic quality control and exclusion of peripheral voxels or voxels covered by fat suppression bands, and automatic highlighting of highest CC/C triplets. Meanwhile, all these steps remain essentially performed by the radiologist. The qualitative approach involves a quite similar evaluation, but with the major exception that all steps related to curve fitting are not needed. Furthermore, many voxels with high citrate and low choline concentration (or vice versa) can be recognized at a glance and do not necessarily need quantification. Therefore, the qualitative approach can speed up the evaluation, with similar diagnostic results as shown in this study.

The overall diagnostic performance parameters of spectroscopy (not combined with morphologic T2-weighted imaging) in our study and its sensitivity for high grade tumours (Gleason 4+3 or higher) are in line with previously published figures [5,19]. For tumour detection, it is of note that discordant findings between the qualitative and quantitative approach always occurred in lower grade tumours. These are indeed frequently associated with citrate and choline peaks of about equal height, resulting in a less obvious assignment to a diagnostic scale (2 or 3) and CC/C-ratios approaching the diagnostic threshold (0.94). On the other hand, high grade tumours usually show lower citrate and higher choline peaks, resulting in more confident diagnostic scale allocation (3 or 4) and higher CC/C-ratios. For the diagnosis of high grade tumours, the choice of diagnostic approach is therefore of minor importance.
The CC/C-ratio cut-off to discriminate cancer from normal tissue in our study (0.94) is higher than in other series (0.48–0.87) [22,29,30]. We believe that this is at least partially due to the quantitative evaluation method that we used. Other series were largely based on selected voxels of good to excellent spectral quality in histopathologically confirmed healthy or malignant tissue (“best case scenario”). We included all voxels of any quality in a list of triplet CC/C-ratios that one by one needed to be evaluated until an acceptable quality was found; by definition, some of them were therefore of borderline quality (“worst case scenario”). This procedure induced a bias towards acceptance of higher CC/C-ratios than would otherwise be done.

The qualitative scoring system and diagnostic results of our study have currently only been tested at a 1.5 T scanner using an endorectal coil. They should not be extrapolated to 3 T scanners, although we do not expect significantly different results due to the higher spectral resolution at 3 T or due to its better discrimination of a polyamine peak (resonating at 3.1 ppm between choline and creatine). Furthermore, our study was a single-reader experience, that now needs confirmation in a multi-reader study to assess any intra- and interobserver variability.

In conclusion, we believe that spectroscopic evaluation with a pattern recognition qualitative approach can be used as a comfortable and reliable alternative to the quantitative approach, with at least equivalent results for prostate cancer detection, especially in high-grade cancers. This might encourage broader acceptance and application of spectroscopy in prostate cancer imaging.

Acknowledgement and disclosure

We gratefully acknowledge Elisabeth Weiland, PhD from Siemens Medical Solutions Germany for making the prototype software (Metabolite Report) available to our department. The authors of this manuscript have no financial conflict with the manufacturer of the used software nor were there any other biases in their scientific work.

References


