Comparison of [¹¹C]choline positron emission tomography with T2- and diffusion-weighted magnetic resonance imaging for delineating malignant intraprostatic lesions


Copyright © 2015 Elsevier Inc. All rights reserved

Published at: http://dx.doi.org/10.1016/j.ijrobp.2015.02.004

PERMISSIONS

http://www.elsevier.com/about/company-information/policies/sharing#preprint

Preprint

- Authors can share their preprint anywhere at any time.
- If accepted for publication, we encourage authors to link from the preprint to their formal publication via its Digital Object Identifier (DOI). Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.
- Authors can update their preprints on arXiv or RePEc with their accepted manuscript.

Please note:

- Cell Press, The Lancet, and some society-owned titles have different preprint policies. Information on these is available on the journal homepage.
- Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles.

9 June, 2017

http://hdl.handle.net/2440/105792
A comparison of $^{11}$C-choline PET with T2 and diffusion-weighted MRI for delineating malignant intraprostatic lesions

Running title: $^{11}$C-choline PET compared with MRI
Abstract

Purpose: To compare the accuracy of $^{11}$C-choline (CHOL) positron emission tomography (PET) with the combination of T2-weighted (T2W) and diffusion-weighted (DW) magnetic resonance imaging (MRI) for delineating malignant intraprostatic lesions (IPLs) for guiding focal therapies and to investigate factors predicting the accuracy of CHOL-PET.

Methods and Materials: This study included 21 patients who underwent CHOL-PET and T2W-/DW-MRI prior to radical prostatectomy. Two observers manually delineated IPL contours for each scan, and automatic IPL contours were generated on CHOL-PET based on varying proportions of the maximum standardized uptake value (SUV). IPLs identified on prostatectomy specimens defined the reference standard contours. The imaging-based contours were compared with the reference standard contours using Dice similarity coefficient (DSC), sensitivity and specificity. Factors that could potentially predict the DSC of the best contouring method were analyzed using linear models.

Results: The best automatic contouring method, SUV$_{60}$, had similar correlations (DSC 0.59) with the manual PET contours (DSC 0.52, $P=0.127$) and significantly better correlations than the manual MRI contours (DSC 0.37, $P<0.001$). The sensitivity and specificity values were 72% and 71% for SUV$_{60}$; 53% and 86% for PET manual contouring; and 28% and 92% for MRI manual contouring. The tumor volume and transition zone pattern could independently predict the accuracy of CHOL-PET.

Conclusions: CHOL-PET is superior to the combination of T2W- and DW-MRI for delineating IPLs. The accuracy of CHOL-PET is insufficient for gland-sparing focal therapies,
however may be accurate enough for focal boost therapies. The transition zone pattern is a
new classification that may predict for how well CHOL-PET delineates IPLs.

Keywords:
11C-choline PET
Diffusion Magnetic Resonance Imaging
Magnetic Resonance Imaging
Prostate cancer
Focal therapy
Introduction

Conventional treatment modalities for localized prostate cancer such as radical prostatectomy and radiotherapy aim to treat the entire prostate gland, with little consideration of the location of the malignant intraprostatic lesion (IPL). This is despite the fact that it is uncommon for localized prostate cancer to involve the entire prostate. One large prostatectomy-based study of predominantly Gleason 7 and pT2c cases found that the tumor involved a median of only 5% of the total prostate volume (1).

Conventional whole-gland treatments pose several problems. Firstly, these treatments may not be necessary in lower risk patients, and may result in unnecessary treatment-related morbidities (2). Secondly, whole gland treatments may not treat IPLs aggressively enough in higher risk patients, resulting in a high likelihood of local recurrences (3-5). Focal treatments using modalities such as intensity-modulated radiotherapy, focal brachytherapy, focused ultrasound and cryotherapy have been proposed as ways of overcoming these problems (2,6-8). Lower risk patients who need more than active surveillance may be treated by focal therapy to treat the IPLs only, while sparing the rest of the prostate gland (2,8,9). Higher risk patients may be treated by whole-gland treatment with a focal boost to IPLs to improve local control (9-12). These focal treatments remain investigational and have not gained widespread acceptance.

One reason why these focal treatments have not been more widely adopted is that conventional imaging with modalities such as T2-weighted (T2W) MRI has not been reliable enough to localize IPLs (2). However, new imaging modalities such as $^{11}$C-choline (CHOL)
positron emission tomography (PET)(13), $^{11}$C-acetate PET(14), diffusion-weighted (DW) and dynamic contrast-enhanced (DCE) MRI(15), and magnetic resonance spectroscopy (MRS)(16,17) may more reliably identify IPLs, potentially making focal therapies a more viable strategy.

CHOL-PET is a particularly promising imaging modality for identifying IPLs. CHOL is a radiotracer based on choline, an essential component of the cell membrane. Prostate cancer cells show changes in choline transport and choline kinase alpha expression, leading to an increased uptake of choline(18). However, CHOL-PET for localizing prostate cancer is currently controversial, with some studies showing high accuracy(13,17), whilst others show no advantages over T2W-MRI(16,19).

The aim of this study is to compare the accuracy of CHOL-PET with T2W-MRI for delineating IPLs based on histopathological reference standards and to investigate factors that can predict the accuracy of CHOL-PET.

Materials and Methods

Study design and patient population

This study is an analysis of data from a prospective, single institution study of 30 patients recruited between September 2008 and March 2011. Eligible patients included those with histopathologically proven prostate adenocarcinoma with intermediate to very high-risk factors(20), who were medically and surgically suitable for radical prostatectomy.
All patients had standard diagnostic and staging investigations prior to recruitment, including serum PSA, trans-rectal ultrasound (TRUS)-guided biopsy, whole body scintigraphic bone scan and CT. Following recruitment, all patients underwent pelvic MRI and pelvic CHOL-PET/CT at least 2 weeks after the TRUS biopsy. DW-MRI was not mandated in the original trial protocol, but was also performed with the standard MRI sequences in 24 patients. Patients underwent radical prostatectomy within four weeks of the MRI and PET scans. Patients who did not have DW-MRI or who had significant imaging artifact that compromised their interpretation were excluded from the final analysis.

All patients provided written informed consent, and the institutional ethics committee approved this study. An analysis of the same patient cohort, investigating Gleason score prediction using imaging has previously been reported(21).

**Imaging protocols**

Detailed descriptions of our CHOL-PET/CT and MRI acquisition protocols have previously been published(13,21). CHOL-PET/CT was performed on an Allegro GSO full-ring 3D PET scanner (Philips Healthcare, Cleveland, OH, USA) and MRI was performed on a 3-Tesla Magnetom Verio system (Siemens, Erlangen, Germany) with an external body-array coil. MRI sequences that were analyzed in this study included multiplanar 2D T2W turbo spin-echo (TSE) which had voxel sizes of 1.9mm x 1.9mm x 4.4mm, 3D T2W sampling perfection with application optimized contrasts using different flip angle evolutions (SPACE) which had voxel sizes of 1.3mm x 1.3mm x 1mm, and apparent diffusion coefficient maps created from DW-MRI which had voxel sizes of 1.2mm x 1.2mm x 4.4 mm. The CHOL-PET scan had voxel sizes of 4mm x 4mm x 4mm.
Pathological specimen preparation

The prostate was step-sectioned at 4mm transverse slices, perpendicular to its posterior surface, as previously described (13). The outline of the tumors was marked directly on the microscope slides of the prostatectomy specimens, and these slides were then scanned directly on a flat bed scanner. A single pathologist reviewed the specimen and outlined each tumor focus on the histologic sections. Marked specimens were scanned directly on a flat-bed scanner.

Image co-registration

The SPACE-MRI sequence was resampled so that the posterior surface of the prostate was vertical, in order to match the slice angle of the prostatectomy specimens. The DW-MRI and CHOL-PET scans were manually co-registered with the SPACE-MRI sequence. The prostatectomy specimen digital images were stacked, and reoriented to match the MRI slices. These images were manually deformed to match anatomical landmarks on the MRI scan as previously described in order to account for shrinkage and distortion of the prostate ex-vivo (13) (Fig. 1).

Generation of reference standard and index test contours
The deformed prostatectomy specimen digital images with IPLs marked by the pathologist were used to delineate the IPLs onto the primary dataset. This was used as the reference standard contour (Fig. 1).

The index test contours consisted of MRI manual contours, CHOL-PET manual contours, and CHOL-PET automatic contours. Two genitourinary radiologists with minimum 12 years expertise (MRI\textsubscript{obs1} and MRI\textsubscript{obs2}) independently delineated the IPLs manually using the combination of multiplanar TSE-, SPACE- and DW-MRI sequences (Fig. 1). Similarly, two nuclear medicine physicians with minimum 7 years expertise (PET\textsubscript{obs1} and PET\textsubscript{obs2}) independently delineated the CHOL-PET IPLs manually. Their contours were designated PET\textsubscript{obs1} and PET\textsubscript{obs2}, respectively. The CHOL-PET automatic contours were generated using the relative standardized uptake value (SUV) thresholding algorithms. Thresholds were set based on the following proportions of the SUV\textsubscript{max} inside the prostate: 40%, 50%, 60%, 70% and 80%. These contours were designated SUV\textsubscript{40}, SUV\textsubscript{50}, SUV\textsubscript{60}, SUV\textsubscript{70} and SUV\textsubscript{80}, respectively (Fig. 1).

**Analysis of correlation**

The degree of correlation was compared by the degree of voxel overlap between the different contours. In order to account for errors in image coregistration, a 5 mm 3D isotropic expansion margin was placed around each contour. Each index test contour was individually compared with the reference standard contours using the following metrics: Dice similarity coefficient (DSC), sensitivity, specificity and Youden index (YI)(13,22,23). These metrics were averaged for each of the two observers for PET and MRI, and designated PET\textsubscript{obsav} and MRI\textsubscript{obsav}, respectively.
The best automatic contouring method was found by determining which one had both the highest DSC and the highest YI. The other automatic contours were not analyzed any further.

The DSC and YI values were compared between PET_{obs}, MRI_{obs}, and the best automatic contouring method using 2-way ANOVA (with patient and contouring method as blocking factors) and least significant difference (LSD) post hoc testing.

**Analysis of factors that may be predictive of the degree of correlation**

The index test contour that had the best correlation with the reference standard contour was investigated further to determine whether or not several patient factors could predict the degree of correlation as determined by DSC. The following factors were investigated: pathological T-stage, initial PSA, prostatectomy Gleason score, percentage pathological tumor volume (tumor volume as a percentage of total prostate volume), SUV_{max}, tumor configuration on prostatectomy specimen and transition zone pattern on axial T2W-MRI. We used the system devised by Souvatzoglou et al.\(^\text{(24)}\) to classify the tumor configurations into four groups: (I) unifocal larger than 5 mm; (II) multifocal; (III) rind-like shaped; and (IV) size < 5 mm. We also developed a system for classifying the transition zone patterns on axial T2W-MRI into four groups: (1) homogenous low signal; (2) other pattern, not otherwise specified; (3) multiple well-defined high signal nodules; and (4) two round, well-defined, heterogeneous regions with low-signal-intensity rim (Fig. 2). Linear models were used to assess the effect of the factors on DSC. Each factor was first fitted independently. A model was then fitted that began with all factors included, and the least significant eliminated at each step until only significant factors remained.
Results

Twenty-one patients of the original cohort of 30 patients were included in the final analysis as 6 patients did not have DW-MRIs and 3 patients had significant susceptibility artifacts (caused by prosthetic hips) on the DW-MRIs. The patient characteristics are summarized in Table 1. The mean DSC, sensitivity, specificity and YI for each contouring method are listed in Table 2. Of the automatic contouring methods, SUV60 had both the highest DSC and the highest YI with values of 0.59 and 0.43, respectively.

The DSCs of SUV60, PETobsav and MRIobsav were 0.59, 0.52 and 0.37, respectively. SUV60 was not significantly higher than PETobsav ($P=0.127$), however was significantly higher than MRIobsav ($P<0.001$). PETobsav was also significantly higher than MRIobsav ($P<0.001$).

The YIs of SUV60, PETobsav and MRIobsav were 0.43, 0.39 and 0.19, respectively. SUV60 was not significantly higher than PETobsav ($P=0.367$), but it was significantly higher than MRIobsav ($P<0.001$). PETobsav was also significantly higher than MRIobsav ($P=0.001$).

SUV60 was found to have the best DSC of all the contouring methods; therefore this method was investigated further to see if any patient factors may be predictive of the DSC. On individual comparisons, the following factors were found to be significantly correlated with DSC: transition zone pattern ($P=0.011$), tumor configuration ($P=0.042$), Gleason score ($P=0.004$), and percentage tumor volume ($P=0.001$). Transition zone pattern 4 had significantly lower DSCs than patterns 1 ($P=0.015$), 2 ($P=0.004$), and 3 ($P=0.005$). Transition zone patterns 1, 2 and 3 were not significantly different from each other. Tumor configuration IV had significantly lower DSCs than configurations I ($P=0.016$), II ($P=0.005$), and III
(P=0.047). Tumor configurations I, II and III were not significantly different from each other. There was moderate positive correlation between DSC and the Gleason score (Spearman’s ρ=0.567, P=0.0074). There was strong positive correlation between the DSC and the percentage tumor volume (Spearman’s ρ=0.665, P=0.001).

On fitting linear models with all of these factors, only the transition zone pattern (P=0.006) and percentage tumor volume (P=0.008) remained significantly correlated with DSC. Patients with transition zone pattern 4 had a mean DSC of 0.35 as compared with a mean DSC of 0.66 for all other patients.

**Discussion**

The main finding in this study is that both manual and automatic contouring using CHOL-PET was superior to manual contouring using MRI for delineating the IPL when correlated with its histopathological reference.

Automatic contouring using CHOL-PET has been assessed in two previous studies. A pilot study of eight patients reported SUV60 as the contouring method with the best correlation with histopathology(13). A similar but smaller study of four patients also reported SUV60 as the contouring method best correlated with histopathology(25).

Several previous studies have compared CHOL-PET with MRI for localizing IPLs. Yamaguchi et al.(17) also found results favoring CHOL-PET as compared with MRI in their study of 20 patients. Other studies found that CHOL-PET was not significantly better than T2W-MRI(16,19). Testa et al.(16) compared CHOL-PET, T2W-MRI, and MRS in 26
patients. The sensitivity and specificity values were 55% and 86% for CHOL-PET, 54% and 75% for MRI, and 81% and 67% for MRS. The best results were obtained with MRS; however CHOL-PET and MRI were similar. Van den Bergh et al. (19) compared CHOL-PET with T2W-MRI in 49 patients. They reported sensitivity and specificity values of 33.5% and 94.6% for MRI scans and 77.4% and 44.9% for CHOL-PET scans. The sensitivity and specificity values of MRI found in our study were similar to Van den Bergh et al’s. The specificity of CHOL-PET using the SUV\textsubscript{60} contouring method in our study was however, significantly higher, with sensitivity and specificity values of 75.0% and 69.7%.

Our study had several significant differences in methodology that may explain the discrepant results with the studies described above. Firstly, all the studies described above performed correlations on the basis of whether or not prostate segments based on imaging matched up with prostate segments based on prostatectomy. The prostate segments used in the above studies were laterality (2 segments), sextants (6 segments) or octants at each of the apex, mid-gland and base (24 segments). Our study performed correlations using voxels (hundreds of voxels per prostate volume) which is a more precise methodology (15). Secondly, the above studies defined PET positive lesions differently. The positive segment was defined either as the one with the highest SUV\textsubscript{max}, subjectively using a nuclear medicine physician, or using varying SUV\textsubscript{max} thresholds. Our study defined positive voxels using both automatic contouring and manual contouring using two independent observers for each scan. The fact that we selected the best contouring method out of a selection of methods together with more refined and objective voxel assessment may have resulted in our study having a better correlation than the previous studies. Differences in patient characteristics and image acquisition protocols may also have contributed to these differences.
The second major finding in this study is that the transition zone pattern can predict for how well CHOL-PET can localize the IPLs. Transition zone pattern 4 is associated with poorer correlations than transition zone patterns 1, 2 and 3 (Fig. 2). Pattern 1 (homogeneous low signal) is usually associated with carcinoma arising from or invading into the transition zone(26,27). Pattern 2 (other pattern, not otherwise specified) is a nonspecific category. Pattern 3 (multiple well-defined high signal nodules) is usually associated with glandular benign prostatic hyperplasia(27). Pattern 4 (two round, well-defined, heterogeneous regions with low-signal rim) is also usually associated with benign nodular hyperplasia(26). Pattern 4 seemed to be associated with intense CHOL-PET uptake within the transition zone, leading to a large false-positive volume, thus leading to poorer correlations. The main implication of this finding is that patients may be more appropriately selected for CHOL-PET scans by first performing a T2W-MRI and stratifying patients according to transition zone patterns. This may be helpful for guiding future clinical trials but will require prospective validation before routine usage.

Souvatzoglou et al.(24) also performed a study to investigate factors that may predict for how well CHOL-PET can localize IPLs. They found that tumor configuration was the only factor significantly negatively influencing tumor prediction. Tumor configuration IV (size of tumor < 5 mm) was associated with significantly poorer correlations than the other tumor configurations. We also investigated Souvatzoglou et al.’s tumor configuration classification system in our study. On an individual comparison, tumor configuration also appeared to predict for the ability to localize IPLs. When all factors were fitted into the linear model (most notably, percentage tumor volume), tumor configuration became a non-significant predictor in our study. The percentage tumor volume may represent a stronger predictor for the ability of CHOL-PET to localize IPLs than tumor configuration.
There are a number of limitations in this study that must be addressed. As with most imaging-pathology correlation studies, there is uncertainty as to the accuracy of co-registration. One source for inaccurate co-registration is a possible discrepancy in slice angles between imaging and pathological specimens. We attempted to account for this by slicing the pathological specimens uniformly and perpendicular to the posterior surface, with reformatting of the imaging to be at the same slice angle for improved correlation. Another source of inaccurate co-registration is the fact that tissue tends to shrink and distort after being removed from the body. We attempted to account for this using deformable registration. An uncertainty expansion margin was also employed, which should compensate for some of the residual errors in registration.

Another limitation lies in the shortcomings of the CHOL-PET scan itself. As shown by Souvatzoglou et al. (24), CHOL-PET cannot distinguish between benign prostatic hypertrophy, prostatitis and prostate cancer. Automatic contours based on the SUV_{max} may therefore be influenced by these benign pathologies. This likely plays a role in limiting the accuracy of SUV60.

Unfortunately, the full set of multiparametric MRI sequences (including DCE-MRI and MRS) was not originally specified as it predated its common usage, therefore was not performed on all patients in this study. Several recent studies have shown multiparametric MRI to be sufficiently accurate to guide focal therapies(6,15), so it would be interesting to compare these with CHOL-PET in future studies.
The main aim of this study was to find a strategy that can be used for guiding focal therapies in prostate cancer. We believe that the methodology employed in this study is appropriate to determine this. Unlike most previous correlation studies, correlations in our study were based on readily reproducible quantification of the overlapping voxels using the different contouring methods. This more accurately reflects the decisions that need to be made by the clinician delivering focal therapies. Decisions need to be made about where to target the focal therapy, which in many cases needs to be within the nearest few millimeters as often IPLs are close to dose limiting normal structures and this assessment is dependant on the imaging methodology used. Since the voxels used in this study are 1-4mm wide, the clinician can be sure that the correlation metrics used in this study are relevant to the treatment. Past studies which performed correlations based on large segments of the prostate (e.g. laterality, sextants, octants) are more relevant for issuing diagnostic imaging reports rather than guiding focal therapies.

The question that then arises is whether or not CHOL-PET automatic contouring is accurate enough to guide focal therapies. For the purposes of guiding focal therapies where the rest of the gland is spared, CHOL-PET automatic contouring may not be sufficiently accurate. With a sensitivity of 72%, 28% of the true IPL volume may still lie outside of the CHOL-PET-defined volume, and therefore may not be treated by these gland-sparing approaches. However, CHOL-PET may be accurate enough to guide focal boost therapy, where the whole gland is treated, and the IPL receives a boost of more aggressive treatment. The accuracy of CHOL-PET reported here is high enough such that focal boost therapy would still significantly increase the therapeutic ratio(9) and warrants full evaluation within the context of a clinical trial.
Conclusion

CHOL-PET is superior to the combination of T2W and DW-MRI for delineating IPLs. The accuracy of CHOL-PET is insufficient for gland-sparing focal therapies, however may be accurate enough for focal boost therapies. The transition zone pattern is a new classification that may predict for how well CHOL-PET delineates IPLs and may be used to more appropriately select patients for this scan.

References


Figure 1. Registration and generation of contours. The T2W-MRI (a) was fused with the DW-MRI (b) and CHOL-PET scan (c). The pathologist provided digital images of prostatectomy specimen axial slices (d), with the involved tumor regions colored in black. These images were then deformed using a mesh to match anatomical landmarks on the MRI scan, and then co-registered with the MRI scan (e). Contours were then created on the MRI scan (f). Four different contours are shown: prostate (green), radiologist-defined manual contour on MRI (blue), SUV$_{60}$ automatic contour on CHOL-PET (yellow), and pathologist-defined tumor (red). Although on average, CHOL-PET contours were more accurate than MRI contours, in some cases (this one included), the MRI contours were more accurate than CHOL-PET contours.
Figure 2. Transition zone patterns. Four different patterns were identified in the prostate transition zones on axial T2W-MRI (1a – 4a). The corresponding fused PET/MRI is shown to the right of these images (1b – 4b). Pattern 1 corresponds to homogeneous low signal (1a). Pattern 2 corresponds to another pattern, not otherwise specified (2a). Pattern 3 corresponds to multiple well-defined high signal nodules (3a). Pattern 4 corresponds to two round, well-defined, heterogeneous regions with low-signal-intensity rim (4a).