The role of MRI and PET/CT in the primary staging of newly diagnosed prostate cancer: a systematic review of the literature

Raisa S. Abrams-Pompe1,2, Stefano Fanti3, Ivo G. Schoots4, Caroline M. Moore5, Baris Turkbey6, Andrew J. Vickers7, Jochen Walz8, Thomas Steuber1, James A. Eastham9

1Martini-Klinik Prostate Cancer Center, University Hospital Hamburg-Eppendorf, Hamburg, Germany
2Department of Urology, University Hospital Hamburg-Eppendorf, Hamburg, Germany
3Department of Nuclear Medicine, Policlinico S. Orsola, University of Bologna, Bologna, Italy
4Department of Radiology & Nuclear Medicine, Erasmus MC University Medical Center, Rotterdam, The Netherlands
5Division of Surgery and Interventional Science, University College London, London, UK & Department of Urology, UCLH NHS Foundation Trust, London, UK
6Molecular Imaging Program, National Cancer Institute, National Institutes of Health, Bethesda, USA
7Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, USA
8Department of Urology, Institut Paoli-Calmettes Cancer Center, Marseille, France
9Urology Service, Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, USA

Word count abstract:
Word count manuscript:

Keywords: Prostate cancer; extraprostatic extension; lymph node metastases; metastases; staging; imaging; magnetic resonance imaging; Positron emission tomography; PSMA-PET/CT

Corresponding authors:
Raisa S. Abrams-Pompe, MD
Department of Urology, Martini-Klinik Prostate Cancer Center
University Hospital Hamburg-Eppendorf
Martinistrasse 52
20246 Hamburg
Germany
rapompe@gmail.com
Abstract

Context: The management of newly diagnosed prostate cancer (PCa) is guided in part by accurate clinical staging. The role of imaging, including magnetic resonance imaging (MRI) and positron emission tomography/computed tomography (PET/CT) in initial staging remains controversial.

Objective: To systematically review studies of MRI and/or PET/CT in the staging of newly diagnosed PCa with respect to tumor (T), nodal (N) and metastasis (M) staging.

Evidence acquisition: We performed a systematic review of the literature using MEDLINE and Web of Science databases between 2012 and 2020 following the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) statement guidelines.

Evidence synthesis: A total of 139 studies (83 on T, 47 on N and 24 on M status) were included. Ninety-nine were retrospective, 39 prospective and one was a randomized controlled trial (RCT). Most studies on T staging examined MRI, while PET/CT was primarily used for N and M staging. Sensitivity for detection of extraprostatic extension, seminal vesicle invasion or lymph node invasion ranged widely. When imaging was incorporated into existing risk tools, gain in accuracy was observed in some studies, although these findings have not been replicated. For M staging, most favorable results were reported for PSMA-PET/CT, which demonstrated significantly better performance than conventional imaging.

Conclusions: A variety of studies on modern imaging techniques for TNM staging in newly diagnosed PCa exist. For T and N staging, reported sensitivity of imaging such as MRI or PET/CT varied widely. The most promising technique for N and M staging, which was recently evaluated in an RCT, was PSMA-PET/CT. Further comparative studies are needed.

Patient summary: We performed a systematic review of currently available imaging modalities to stage newly diagnosed PCa. With respect to local tumor and lymph node assessment, performance of imaging ranged widely. However, PSMA-PET/CT showed favorable results for detection of distant metastases.
1. Introduction
While localized prostate cancer (PCa) is curable using surgery or radiation therapy, cure is unlikely in the presence of metastatic disease, despite new systemic treatments, which have improved survival in the metastatic setting. (2-4). Therefore, appropriate assessment of the extent of PCa at diagnosis is critical in guiding initial treatment.

Current guidelines recommend abdominopelvic imaging in men with intermediate- and high-risk disease at risk of lymph node metastases, as well as bone scintigraphy using 99mTc-labeled bisphosphonates in selected men with intermediate- as well as in all men with high-risk disease. (1, 2) Unfortunately, conventional imaging with computed tomography (CT) and bone scintigraphy suffer from a lack of sensitivity and specificity in identifying metastatic cancer, which has prompted the search for new imaging techniques with better diagnostic accuracy.(3) For local tumor and lymph node staging, multiparametric magnetic resonance imaging (mpMRI) has gained more and more attention. In 2012, the European Society of Urogenital Radiology (ESUR) standardized MRI reporting by introducing the Prostate Imaging Reporting and Data System (PI-RADS). (4) In 2015, this version was updated in collaboration with the American College of Radiology (ACR) to PIRADS v2.(5)

Positron emission tomography/computed tomography (PET/CT), initially with radiolabeled fluorodeoxyglucose (FDG), was evaluated; however, for relatively well-differentiated tumors, FDG-PET/CT was not beneficial.(6, 7) Newer tracers, such as 18F-Sodiumfluoride- (NaF-), 18F/-11C-choline, 18F-fluciclovine (FACBC), and 68Ga-labelled PSMA have recently been developed and evaluated. With high sensitivity and specificity, 68Ga-labelled PSMA PET/CT has been rapidly adopted. While the majority of published articles examined PET/CT in the context of recurrent PCa, there are only few studies analyzing staging modalities in newly diagnosed PCa.(8) This prompted us to perform a systematic review of the current literature on modern imaging for staging of newly diagnosed PCa.
2. Evidence Acquisition

2.1 Research question
The aim of this systematic review was to examine the role of modern imaging types, such as PET/CT, PET/MRI, mp/whole body-MRI for staging (TNM) of newly diagnosed PCa and to report their imaging test accuracy.

2.2 Search strategy
We performed a systematic review of the literature using MEDLINE and Web of Science databases between 2012 and 2020 following the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) statement guidelines.(9) The following search strategy was used as keywords and/or free texts: (“prostate cancer” OR “prostate neoplasm”) AND (“MRI” OR “PET CT”) AND (“staging” OR “tumor stage” OR “lymph nod*” OR “metastas*”). Furthermore, cited references from selected articles and from review articles retrieved in the search were screened for additional information. All abstracts were screened by two independent reviewers (RSAP) and (JE) using a newly developed standardized data form. Any disagreements were resolved by open discussion. Based on title and abstract selection, full texts were analyzed in detail for eligibility for the final review. The validated Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) scoring system was used to assess the risk of bias.(10)

2.3 Study selection
Inclusion criteria followed the Patient Index test Comparator Outcomes Study (PICOS) design: participants, interventions, comparisons, outcomes, and study design. Therefore, studies were assessed considering patients with biopsy-proven, newly diagnosed PCa (P) who underwent MRI and/or PET CT (I) for further disease assessment (O) with respect to tumor stage, lymph node status and distant metastases. Only original articles and brief correspondences were included (S). Most studies did not have a comparator (C) group. In some, conventional imaging such as contrast CT or bone scintigraphy served as comparator. In some, MRI was directly compared to PET/CT. For tumor (T) and lymph node (N) staging, only studies with histological confirmation by radical prostatectomy (RP) or pelvic lymph node dissection (PLND) as “gold” standard of reference were included. T stage included extent of tumor beyond the capsule (≥pT3a/b), while studies analyzing index tumor,
tumor detection or localization of primary tumor were not considered. For metastases (M) staging, a best value comparator mostly derived by panel decision considering clinical, biochemical and imaging data at baseline as well as at follow-up was used. In some studies, M status was additionally confirmed by histology, e.g. via bone biopsy. Studies without either a best value comparator or histological confirmation were excluded. The search was limited to English-language articles. We also included articles, in which results of subgroups with primary PCa staging were reported separately. However, articles presenting mixed results of staging and restaging purposes were not considered.

2.3 Data extraction
From each selected study, we extracted first author, year of publication, study design, imaging type, imaging technique (including tracer or sequences), total number of patients analyzed, main patient characteristics, endpoint and detection rate, number of readers, sensitivity, specificity, negative (NPV) and positive predictive value (PPV) as well as accuracy. For M staging, in which standard of reference was other than histological confirmation, standard of reference was extracted. Whenever possible, sensitivity, specificity, NPV and PPV were calculated. If performance was assessed on region and patient basis, we included patient-based results.

3. Evidence Synthesis
The heterogeneity of the studies entailed that summary statistics from different studies could not be combined meta-analytically. Hence we summarized the results narratively.

3.1. Characteristics of included articles
Between 2012 and 2020, 4170 studies were identified using our search criteria (Figure 1, PRISMA flow diagram). Title and abstract screening resulted in 360 studies that entered full-text assessment. An additional 25 studies were retrieved through reference screening. After full-text assessment, another 246 studies were excluded. Thus, 139 studies remained eligible for inclusion in this review. Of these, 83 examined imaging for T, 47 for N and 24 for M staging of newly diagnosed PCa. Fifteen studies reported on several endpoints, while most commonly T and N stage were combined. Thirty-three studies compared different imaging modalities. In 13
reports, imaging modalities were compared or incorporated into currently used nomograms such as the MSKCC or Briganti nomograms, or the Partin Tables. Most patient cohorts, especially for N and M assessment consisted of intermediate and/or high-risk patients while for T staging, several studies also included low- or favorable intermediate-risk patients. Sample size varied widely, ranging from 10 to 1045 included patients. The study design comprised primarily retrospective series (71%). However, 40 studies reported on prospective series and there was one randomized controlled trial (RCT).

### 3.2 Quality of studies

Quality of included studies differed widely and was overall moderate (Supplementary Tables 1 – 3, Supplementary Figure 1). For T and N status risk of bias was rated lower compared to M status. Regarding patient selection risk of bias was rated unclear in most studies, as patient enrollment was not reported. All studies for T and N staging reported on pre-selected patients, as only RP candidates were included. For M staging, most cohorts consisted of high-risk patients. Index test was considered low risk of bias if readers were blinded to clinical data and reference standard results and if interpretation was done in a standardized fashion. Although many studies reported blinding, interpretation differed among readers and many lacked standardized reporting. Interpretation of reference of standard was considered at low risk of bias in case of histological confirmation and blinding to index test results. Regarding T and N status, only studies with histological confirmation by RP and/or PLND were included. Therefore, risk of bias would have judged low for T and N stage. However, most studies lacked information on blinding to index test results. For M status all studies were found to have high risk of bias as reference standard consisted of a best value comparator using different definitions and follow-up periods. Moreover, follow-up imaging was interpreted with knowledge of index test results and therefore inevitably at high risk of bias. Flow and timing was rated unclear in most studies, as time interval between index tests and/or standard reference was not reported. Concerns of applicability were present in only few studies. In some, imaging techniques or interpretations varied as imaging was performed in external centers and interpreted by different readers without standardized reporting.
3.3. T stage: detection of extraprostatic extension and seminal vesicle invasion

We identified 43 studies that examined the role of imaging for extraprostatic extension (EPE), 31 for seminal vesicle invasion (SVI), and 33 for overall presence of ≥pT3 disease. Table 1 summarizes the main studies reporting on T staging in primary PCa. A total of 77 studies examined performance of MRI (59 mpMRI) and eleven of PSMA PET/CT or PET/MRI including five studies that compared MRI to PSMA–PET. In addition, one study compared results of mpMRI to 18F-Fluorocholine-(FCH-) PET/CT. While study design was retrospective in 67 (81%), 15 studies reported prospective results. Sample size ranged widely from 21 to 1045 included patients. Study populations were notably heterogeneous; however, most frequently mean/median PSA was ≤10ng/ml and Gleason score (GS) ≤7. Six studies focused on higher-risk patients. Furthermore, MRI techniques, e.g. 1.5 or 3T, use of an endorectal coil, including diffusion-weighted imaging, were different among included studies and most studies included MRIs that were performed pre- and post-biopsy. In addition, definition of EPE varied between studies, considering focal EPE in some, while others defined EPE as established.

3.3.1 mpMRI

Current guidelines recommend mpMRI, which combines morphological T2 weighted with functional imaging sequences (diffusion weighted, dynamic contrast enhanced) for pre-biopsy assessment. A remarkable number of studies examined the role of MRI or mpMRI in the context of local T staging. Sample size and event rate among included studies and thus confidence intervals differed widely. Unsurprisingly, sensitivity varied enormously (0 – 100%) between different studies on MRI for detection of EPE and/or SVI (Table 1). However, some findings are worth reporting in more detail. The largest series to date was published in 2017 by Lee et al., including 1045 patients who underwent mpMRI before RP at a single institution. EPE was noted in 314 (27%) patients. Although mpMRI were reviewed by only two experienced radiologists, blinded to all clinical data, sensitivity and specificity in this retrospective study remained relatively low (53% and 82%). It is of note that different MRI techniques (1.5-3T) and no standardized reporting were used. Moreover, most patients had GS 6 (48%) or 7 (36%) and median PSA was 6.1ng/ml, representing a lower risk cohort with presumably lower rates of EPE or SVI.
In that regard, several studies thought to assess performance of MRI for different risk groups.\(11, 22, 23, 27, 41, 43, 49, 62, 65, 69, 73, 78\) While better performance for high-risk patients has been presumed in literature, the data actually remain inconclusive. For example, Jeong et al. analyzed 922 high-risk patients undergoing 1.5-3T mpMRI and reported sensitivity for EPE and SVI of only 43% and 35%.\(42\) In this study, EPE was noted in 530 (58%) and SVI in 117 (13%) men. MRIs included only partly diffusion-weighted imaging and no standardized reporting were used. Moreover, Jansen et al. observed comparable and moderate sensitivity for prediction of EPE between 133 high-risk and 297 low-risk patients (49% vs. 42%, \(p=0.5\)).\(41\) In this study, mpMRIs were evaluated by different radiologists applying standardized reporting in approximately 60% of cases. Using PIRADSv2, Alessi et al. reported considerably high overall sensitivity of 99% and a small increase between 137 low- and 164 intermediate/high-risk patients (96% vs. 100%), although this was not statistically significant.\(11\) One study assessed performance of mpMRI with respect to race. Falagario et al. reported on 975 African and Caucasian American (CA) men undergoing preoperative mpMRI and subsequent RP.\(27\) A stage \(\geq pT3\) was noted in 255 patients (26%). While there was no difference with respect to race, sensitivity of MRI was lower for low-risk compared to high-risk Caucasian Americans (28% vs. 58%).

Interpretation of mpMRI might vary between different readers with an assumed benefit for radiologists with high level of MRI experience.\(15, 33, 61, 68, 74, 76, 80\) Though most studies reported better performance among radiologists with high level of experience, most studies consisted of only small sample size with wide and overlapping confidence intervals. Using logistic regression models, Tay et al. observed only small incremental benefit of mpMRI over clinical parameters when standard radiological reports were considered.\(80\) However, EPE classification increased significantly by adding a specialized report of a dedicated expert in prostate mpMRI (AUC 0.91 vs. 0.69, \(p<0.001\)). At closer examination, this difference was due to improvements in specificity (44% vs. 81%) while sensitivity remained comparable (77% vs. 86%).

To overcome variability due to subjective interpretation, the ESUR has introduced a standardized reporting system for mpMRI (PIRADSv2, ESUR EPE score). Schieda et al. analyzed performance of mpMRI for predicting EPE in 145 men with respect to the use of PIRADS vs. no PIRADS classification.\(76\) The authors observed
significantly better sensitivity and accuracy for EPE among experienced radiologists without the use of standardized reporting; interestingly, this difference disappeared when PIRADS was applied. Furthermore, overall accuracy increased with the use of PIRADS (42% vs. 63%, p=0.006). Again here, sample size was relatively small. Kam et al. observed significant improvement in sensitivity for prediction of EPE from 30% to 60% when applying PIRADS v2 compared to v1 in 235 patients (p=0.008). However, this study lacked information on patient (tumor characteristics, MRI findings, event rate, confidence intervals) or further outcomes for the different groups. Overall, interobserver agreement was poor to moderate in most studies and moderate to good for studies using PIRADS v2.(39, 53, 56, 61, 68, 74, 75)

Another approach to increase sensitivity, usually at a cost of specificity, consists of combination of indirect and direct MRI signs of EPE, although studies remain too small for firm conclusions to be drawn.(19, 29, 59, 84, 88)

Four studies examined the performance of MRI when performed and interpreted in non-academic settings.(13, 22, 51, 55) Davis et al. assessed performance of mpMRI performed in community centers among 133 patients and reported a sensitivity of 0% in a subgroup of 52 low-risk patients.(22) Lebacle et al. showed sensitivity of 35% in a cohort of 853 patients that underwent MRI externally without any restrictions.(51) However, as up to date no studies are available that directly compare quality and performance of MRIs performed in academic vs. non-academic settings

Several MRI factors other than direct indications of EPE have been proposed to predict EPE such as tumor contact length, tumor diameter, primary lesion score or size.(31, 48, 53, 72, 92) No external validation of these factors has been conducted to date.

3.3.2 PSMA-PE/CT or PET/MRI

PET/CT or PET/MRI using radiolabeled tracers such as choline have shown good sensitivity, although limited specificity, for detecting prostate cancer. Prostate-specific membrane antigen, PSMA-PET/CT is thought to overcome this issue and has demonstrated its value in the context of recurrent PCa.(8) For staging at diagnosis, most studies included intermediate- or high-risk patients. In these, PSMA-PET/CT or PET/MRI offers the advantage of whole body imaging combined with local staging, which might result in lower costs and time saving, than pelvic MRI. Eleven studies reported on the use of 68Ga-PSMA PET/CT or PET/MRI for local tumor staging of
primary PCa and described mixed results. Most studies included only small patient cohorts or had few events, inherent with high level of uncertainty, resulting in a wide range of reported sensitivity ranging from 0% to 94% for EPE as well as from 11% to 94% for SVI.(12, 24, 28, 34, 35, 57, 58, 81, 85, 86, 90) For example, high sensitivity of 94% for prediction of EPE and 83% for SVI was found in the study by Thalgott et al. that assessed PSMA-PET/mpMRI for local tumor staging in 73 high-risk patients, including 53 (71%) with EPE as well as 33 (45%) with SVI.(81) Sensitivity of 46% for SVI was observed in the study by Van Leeuwen et al. analyzing 140 men of which 43 (31%) presented with SVI.(85) Conversely, in a smaller study by Dekalo et al., including 59 intermediate- and high-risk patients, sensitivity for SVI was 58% while PSMA-PET/CT detected none out of 17 patients with EPE.(24)

### 3.3.3 Comparisons and incorporation into clinical risk tools
The most commonly used clinical risk stratification tools relying on clinical parameters and biopsy results to predict EPE are the Partin Tables and the MSKCC nomogram.(95, 96) A comparison or incorporation into these staging tools was performed in ten studies.(22, 24, 26, 30, 33, 36, 48, 81, 91, 92) All but two reported better performance for imaging (mostly MRI) although external validation is pending. Highest accuracy was achieved when imaging was incorporated into existing models. For example, Gupta et al. observed AUC of 0.82 vs. 0.62 for a model using mpMRI to predict EPE compared to Partin Tables.(36) Thalgott et al. reported superior sensitivity for imaging (PSMA PET/MRI) over MSKCC nomogram or Partin tables for prediction of EPE (94% vs. 66% vs. 71%).(81) However, accuracy for EPE and SVI did not differ between MSKCC nomogram and imaging. Incorporating mpMRI into Partin Tables, an AUC of 0.93 was reported, incorporation mpMRI into MSKCC, AUC of 0.95 was achieved.(30) Similar gain in AUC was observed for incorporation of PSMA-PET into the MSKCC nomogram (0.84 to 0.91).(24) Five studies thought to compare mpMRI to PSMA-PET/CT.(12, 57, 58) However, none of the study could demonstrate significant superiority of one modality to another.

### 3.4 N stage: Detection of lymph node metastases
Table 2 summarizes studies reporting on imaging for N staging. Similar to local tumor staging, due to widely heterogeneous studies, we found a wide range of sensitivity
and specificity from 10 – 100% and from 33 –100%, respectively. A total of 17 studies were prospective, while 30 reported retrospective results. Most studies relied on intermediate- or high-risk patients and in 13 different imaging modalities were compared.

3.4.1 PET/CT
Hybrid PET/CT or PET/MRI combines the advantages of conventional CT/MRI with PET, resulting in a combination of morphological and anatomical information derived by CT/MRI with additional functional (metabolic/biochemical activity) information provided by PET. By using MRI instead of CT, ionization radiation can be spared. While PET/CT is already widely adopted within staging of recurrent PCa, only few studies reported on its role for primary staging.(8)

3.4.1.1 PSMA-PET/CT
Introduced in 2012, the 68Ga-labelled PSMA-targeted radio-ligand Glu-NH-CO-NH-Lys-[68Ga-(HBED-CC)] (68Ga-PSMA-HBED-CC or 68Ga-PSMA-11) revolutionized PCa imaging. PSMA, a large extracellular type-2 transmembrane glycoprotein, is highly overexpressed in PCa and can easily be targeted by this ligand for imaging purposes.(97) Twenty-four studies examined performance of PSMA-PET for N staging and reported an overall high specificity of 80 – 100%, while widely varying sensitivity of 33 – 100%. (12, 24, 34, 35, 81, 85, 90, 98-114) However, most studies were limited by small patient sample and low event rates with accordingly large confidence intervals, ranging in some between 0 to 100%. Besides study design, size of lymph node metastases (LNM) was a limiting factor. Although PSMA-PET is thought to perform better than conventional imaging, which is based on morphological signs, size of metastatic lymph nodes was noted in most studies as an important limitation with correctly identified LNM to be somewhere around ≥10mm of size. (34, 35, 98, 103, 107, 108, 111-115) Yaxely et al. reported sensitivity of only 38% and specificity 94% among 208 intermediate- to high-risk patients.(108) PSMA-PET/CTs were evaluated by experienced nuclear physician radiologists. Histopathological examination revealed LNM in 55 men (26%). In this study, PSMA-PET/CT correctly identified only 15% of LNM that were <5mm of size. Furthermore, Maurer et al. showed sensitivity and specificity of PSMA-PET/CT to detect LNM of 66% and 99% while accuracy reached 89% in a cohort of 130 men, including 41
with LNM. Maximum size of missed LNM by PSMA-PET/CT was 3mm (1-5mm). Zhang et al. reported sensitivity of 93% and specificity of 96% among 42 men including 15 (36%) with LNM.(114) However, sample size remained relatively small and notably, >80% of all LNM in this study were >10mm in size.

PSMA-PET might not perform inferior to existing prediction tools such as the MSKCC or Briganti nomograms or the Partin Tables.(24, 81, 100) Thalgott et al. found the largest AUC for PSMA-PET/CT (0.8) but this was not statistically different to AUC obtained with the MSKCC nomogram (0.77) or Partin Tables (0.67).(81) A model integrating information of PSMA-PET/CT and the MSKCC achieved significant gain in AUC (to 0.87). However, external validation has not been not performed yet. Furthermore, including quantitative PET parameters such as $\text{SUV}_{\text{max}}$, PSMA$_{\text{vol}}$ might improve accuracy.(100) Likewise, this has to be confirmed in further studies.

Six studies performed comparisons between MRI and PSMA-PET/CT or PET/MRI for N staging.(85, 90, 111-114) Results remained inconclusive as most studies contained only few patients (N= 10 – 42) and reported mixed results. Of particular note is the study by Leeuwen et al. that observed better sensitivity for PSMA-PET/CT compared to 1.5T mpMRI (53% vs. 14%) in a cohort of 140 men including 51 (36%) with LNM.(85) The smaller study by Zhang et al. reported similar performance of high resolution, 3T mpMRI vs. PSMA-PET/CT in detection of LNM (sensitivity of 93% and specificity of 96% for both).(114) However, as mentioned above, this study included a high proportion of LNM >10mm, which might have contributed to the more favorable results.

3.4.1.2 $^{11}$C-Choline-PET/CT

As a phosphatidylcholine, $^{11}$C-Choline is part of cellular membranes and has less urinary excretion than other choline derivates such as FCH resulting in favorable tumor-to-background ratio.(116) In this systematic review, a total of five papers were found that reported on $^{11}$C-Choline for primary N staging. Sensitivity of per patient-based analyses ranged between 10% and 70% and specificity between 76% and 100%.(110, 117-120) As previously reported studies were highly heterogeneous with respect to sample size, patient characteristics or number of examined LN contributing to this range. Three studies compared $^{11}$C-Choline-PET/CT to DW-MRI and reported non-inferior performances although studies were limited by small sample size.(117-119) Interestingly, Vag et al. thought to define optimal ADC and $\text{SUV}_{\text{mean}}$ cutoff
values for prediction of LNM. Highest sensitivity and accuracy for 11C-Choline PET/CT and DW-MRI were observed for SUV$_{\text{mean}}$ threshold of 2.5 and ADC of $1.01 \times 10^{-3} \text{mm}^2/\text{s}$. However, the study included only 34 intermediate- and high-risk patients and findings need confirmation in future studies.\(^{(117)}\) Only one small study directly compared twelve patients undergoing 11C-choline-PET/MRI to twelve patients with PSMA-PET/MRI and reported similar sensitivity, but however higher accuracy for PSMA-PET/MRI.\(^{(110)}\)

3.4.1.3 11C-Acetate-PET/CT
Similar to 11C-Choline, 11C-Acetate offers the advantage of minimal urinary excretion with the benefit of low background radioactivity.\(^{(121)}\) Only three reports on the use of 11C-Acetate-PET/CT for N staging were identified.\(^{(122-124)}\) All were prospective, including 9 to 102 patients. The largest series by Haseebuddin et al., analyzed 102 patients with preoperative 11C-Acetate-PET/CT before RP.\(^{(123)}\) LNM were notice in 21 (21\%) patients. Sensitivity and specificity were 68\% and 78\%. PET positive findings were a significant predictor for treatment failure in multivariable analysis. Interestingly, patients with false positive findings had worse treatment-failure free survival rates compared to patients with true negative results.

3.4.1.4 18F-Fluorocholine
The PET tracer FCH has considerably longer half-life compared to 11C-Choline.\(^{(125)}\) However, urinary excretion remains substantially higher. We found five studies on FCH-PET/CT in primary PCa staging (Table 2).\(^{(63, 125-128)}\) With limitations analogues to previous modalities, reported sensitivity of FCH-PET/CT ranged between 10\% and 78\% and specificity between 69\% and 100\%. Poulsen et al. reported sensitivity of 73\% among 210 patients including 41 (20\%) with LNM. Median number of LN removed was five and therefore relatively low. Mean diameter of true positive nodes was 10.3mm and therefore significantly larger compared to mean diameter of true negative nodes (4.6mm).\(^{(128)}\) Only one study compared FCH-PET/CT to DW-MRI and concluded that performance of FCH-PET/CT might be superior.\(^{(63)}\) However, this study included only 47 patients with as few as nine having LNM.\(^{(63)}\)

3.4.1.5 18F-Fluciclovine-PET/CT
FACBC has already demonstrated its value in the setting of biochemical recurrence and was therefore approved by the US FDA for detection of recurrent PCa in 2016.(129) Within three small studies, including 26–28 patients sensitivity and specificity of FACBC-PET/CT or PET/MRI ranged between 14 – 76% and 86 – 100%.(40, 130, 131) Consistent with reports on other PET tracers, one of the main limitations of FACBC-PET/CT relies in LN size with inability to detect LNM below 7-8mm.(131) Only one out of seven patients was correctly identified in the study by Jambor et al., reporting on 26 patients that underwent FACBC-PET/CT and PET/MRI in a single center.(40) Median size of missed LNM was <8mm. Selnaes et al. compared FACBC-PET/MRI results to 3T mpMRI and reported similar sensitivity of 40% but higher specificity for FACBC-PET/MRI compared to mpMRI in 28 patients, including 10 (38%) with LNM.(131)

3.4.2 mpMRI
A total of 19 studies reported on MRI for N staging in primary PCa. Similar to the results observed for local staging, there was large variation in patient sample and MRI techniques, which resulting wide range of sensitivity from 14% to 100%.(13, 25, 42, 63, 85, 90, 109, 111, 113, 114, 117-119, 131-136) With the exception of some small case series, specificity remained high within the majority of studies (Table 2). Most studies examined the role of diffusion-weighted MRI (DW-MRI), which offers the advantage of imaging without need for exogenous radiolabelled tracers or contrast agents. Similar to previous modalities, several studies highlighted the importance of LNM size.(63, 111, 113, 118, 135) Usually, LN are assumed to be suspicious on MRI with short axis of >10mm in oval or >8mm in round shaped LN. In this review, size of truly detected LNM was somewhat >10mm.(63, 135) Some articles suggested other parameters of mpMRI such as PIRADS lesion score or apparent diffusion coefficient (ADC) values to be more accurate in predicting LNM than size.(111, 117, 132, 134)
For example, Brembilla et al. analyzed 101 patients with risk for LNM of >10% on Briganti nomogram and reported sensitivity for detection of LNM of 91% for presence of PIRADS ≥4 lesions or tumor volume ≥1cc compared to only 17% and 33% sensitivity for presence of enlarged LN or restricted diffusion LN.(132) In twelve studies, MRI was compared to PET/CT scans.(63, 85, 90, 109, 111, 113, 114, 117-119, 131, 133) Results were reported within the different PET/CT sections and Table 2.
3.5 M stage: Detection of metastatic disease spread

Evaluation of distant metastases remains challenging due to the absence of a histological “gold standard”. Therefore, studies using a best value comparator, consisting of consensus considering all available clinical, biochemical and imaging information at baseline and/or follow-up were included in this review. Overall, 24 studies examined imaging for M stage and reported sensitivity of 80 – 100%. Ten studies were prospective and this section includes results of the first RCT on PSMA-PET/CT.(137)

3.5.1.1 PSMA-PET/CT

Six studies were found to evaluate the role of PSMA-PET/CT in primary staging of PCa and reported overall favorable sensitivity and specificity for detection of bone metastases.(137-142) In all studies, PSMA-PET/CT was compared and outperformed either conventional imaging (BS +/- CT/MRI), single photon emission computed tomography (SPECT) or other modern imaging modalities such as NaF-PET/CT or whole body MRI (WB-MRI). The largest and most recent report represents the only randomized controlled trial (RCT) in this setting. Within the proPSMA trial, Hofman et al. reported the results of a multicenter, two-arm randomized study comparing PSMA-PET/CT to conventional imaging consisting of contrast enhanced CT and BS with SPECT-CT.(137) A total of 302 patients with high-risk characteristics underwent randomization. The authors observed a significantly higher accuracy for PSMA-PET compared to conventional staging (92% vs. 65%) for the entire cohort (N and M staging) as well as for distant metastases (95% vs. 74%). Sensitivity and specificity in detection of metastatic disease were 92% and 99% compared to 54% and 93%, respectively for conventional imaging. Interestingly, patients undergoing conventional imaging exhibited 10.9 mSv higher radiation exposure than PSMA-PET/CT. Moreover, less equivocal findings were described using PSMA-PET/CT compared to conventional imaging resulting in reduction of further investigations, which are often needed in case of inconclusive findings. Furthermore, two prospective and three retrospective studies with sensitivities ranging between 96 – 100% and accuracy of 95 – 100% were found. Lengana et al. reported prospectively on a cohort of 113 patients undergoing PSMA-PET/CT and bone scintigraphy for detection of bone metastases.(140) With an
overall detection rate of 25/26 patients with bone metastases, sensitivity and specificity for PSMA-PET/CT were favorable with 96% and 100%. One study thought to compare WB-MRI, PSMA- and NaF-PET/CT and reported significantly higher accuracy for PSMA-PET/CT than WB-MRI, while there was no statistical difference between PSMA- and NaF-PET/CT.(142) However, this study was limited by the small and inhomogeneous study population, including only ten patients for staging purposes, three under Active Surveillance/Watchful Waiting and 37 under ADT.(142)

3.5.1.2 18F-Sodiumfluoride-PET/CT
Another promising bone-specific radiopharmaceutical in the assessment of bone metastases is NaF. NaF binds to mainly osteoblastic bone lesions and - in combination with CT – may offer whole body examination.(143) A total of six studies, including 37 – 211 patients, assessed the use of NaF-PET/CT in detection of bone metastases and reported overall favorable sensitivity from 88 – 100%.(142, 144-148) Two studies consisted of mixed cohorts, including restaging of patients who were already under treatment.(142, 147) Consistent in all studies, NaF-PET/CT performed better than conventional imaging and had less equivocal findings that require additional imaging for further clarification.(144, 146, 148) Zacho et al. reported high interobserver agreement in the detection of bone metastases of two well trained radiologists (Cohen’s kappa 0.89).(147) Poulsen et al. compared NaF- and FCH-PET/CT to bone scintigraphy for detection of spine metastases and reported similar performance of NaF- and FCH-PET/CT while superior performance of NAF-PET/CT compared to bone scintigraphy.(148) However, the study included only pre-selected patients with bone metastases and might not be applicable to other patients and locations.

3.5.1.3 FCH-PET/CT
Within a total of six studies, including 18 – 143 patients, examined the performance of FCH-PET/CT. Reported sensitivity and specificity ranged from 80 – 100% and 91 – 100%, respectively.(126, 127, 148-151) Comparison of FCH-PET/CT to other imaging modalities including conventional imaging, WB-MRI or NaF-PET/CT was performed in five studies.(127, 148-151) While FCH-PT/CT was declared to perform better than conventional imaging, most studies included only few patients and had low event rates.(127, 149, 150) Metser et al. compared FCH-PET/CT or PET/MRI to
WB-MRI (n=48) and did not find a statistically significant difference with respect to skeletal metastases while the authors observed an advantage for FCH-PET/CT in the detection of non-regional lymph node metastases.(149)

3.5.1.3 other PET/CT tracers
Three studies reported on other imaging modalities such as 18FDG-, 11Acetate- or 13N-Ammonia- PET/CT and reported promising results.(133, 152, 153) Two studies assessed use of FDG-PET/CT for primary staging. While for relatively well-differentiated tumors, FDG-PET/CT remained less useful, favorable sensitivity for FDG-PET/CT ranging between 90 – 100% in two cohorts of high-risk patients was reported.(6, 7, 133, 153) When compared to 13N-Ammonia, both tracers had perfect sensitivity for detection of bone metastases.(153) One study reported favorable sensitivity and specificity for 11-Acetate-PET/CT versus bone scintigraphy in detection of bone metastases (100% vs. 69% and 98% vs. 94%).(152) However, all studies were limited by small patient number, low event rate and preselected patients that hinder final conclusions.

3.5.2 WB-MRI
A total of six studies reported on WB-MRI and additional three on pelvic MRI for M staging in newly diagnosed prostate cancer.(142, 145, 149, 151, 154-158) Sensitivity and specificity to predict bone metastases ranged between 74 – 100% and 83 – 100%, respectively, for WB-MRI as well as 71 – 95% and 95 – 100%, respectively, for pelvic MRI. However, study populations and event rates were highly heterogeneous resulting in extremely wide confidence intervals in some studies. Pasoglu et al. prospectively combined mpMRI and WB-MRI as a “one-step TNM staging” for detection of bone metastases in 30 high-risk patients.(154) Both non-irradiation imaging modalities were done within less than one hour during a single visit. Sensitivity and specificity were perfect (100%), though only nine patients had bone metastases. Within a second analysis, the authors postulated higher signal-to-noise ratio and contrast-to-noise ratio for 3D T1-weighted sequences compared to 2D sequences.(155) While this resulted in performance benefit for 3D with respect to N staging, both sequences performed equally well for bone metastases. Eyrich et al. analyzed more than 600 primary prostate cancer patients across all risk stages among 44 different academic and community practices that underwent mpMRI
(pelvis to aortic bifurcation) in addition to conventional bone scintigraphy. Depending on mpMRI interpretation (including equivocal signs), performance was inferior or equal compared to bone scintigraphy. Four studies compared WB-MRI to other modern imaging modalities (PSMA-, NaF- and FCH-PET/CT). Mosavi et al. reported favorable results for WB-MRI and NaF-PET/CT in 49 high-risk patients (100% sensitivity for both). However, only five patients out of 49 patients had bone metastases. Likewise, Metser et al. reported similar performance of FCH-PET/CT and WB-MRI in detection of bone metastases (see 3.5.1.3).

4. Discussion

The aim of this systematic review was to provide an overview on modern imaging modalities for TNM staging of newly diagnosed prostate cancer. We identified a variety of studies and different imaging modalities, especially with respect to N and M staging. Most studies assessing local tumor stage reported on the use of mpMRI, which has gained more and more attention within the last decade. In the latest update of the EAU guidelines, there is a strong recommendation for the use of mpMRI in the pre-biopsy setting. However, no such recommendation for use of mpMRI in further T or N staging exists. The “gold” standard for N staging represents standard or extended PLND, which causes morbidity and may miss lymph node metastases outside the field. Identification of lymph node metastases for further treatment planning, especially in patients that do not undergo RP remains challenging. Compared to pelvic MRI, PET/CT offers the benefit of combined whole-body examination, resulting in detection of lymph node metastases outside the pelvic area.

In this review, sensitivity and specificity of modern imaging for T and N staging ranged from 0% to 100%; in short, its properties are unknown. The wide range of reported performance reflects the heterogeneity of included studies. Most studies were limited by insufficient sample size and event rate with accordingly high level of uncertainty. In addition, 95% confidence intervals were often missing and in some studies, confidence intervals would range considerably wide around reported rates. Moreover, differences in study populations, histopathological interpretations, evaluation methods, imaging technology and reader experience might contribute to
this wide range. Therefore, comparisons of reported results and assessments of clinical significance have to be made with caution.

Another explanation for the wide range of reported sensitivity was due to various definitions of outcome variables. For example, some studies included microscopic lesions for EPE definition while others did not. Level of radiological experience might be of importance and absence of central radiologic review for studies reporting on imaging in external centers might also contribute to the unsatisfying results. Although, the ESUR tried to standardize reporting by introducing PIRADS and an EPE score, presence of EPE/SVI and lymph node metastases still reflects subjective interpretation.

Possibly, imaging might improve local staging when combined with other clinical data and the incorporation into existing risk stratifications such as the MSKCC nomogram or Partin Tables. However, there are no external validations of those models so far. As there are only a handful of studies comparing different imaging modalities, at this moment, we cannot comment on superiority of one modality to another.

Most studies on T staging reported on MRI, while the clinical utility of other imaging modalities such as FACBC- or PSMA-PET/CT remains unknown. Table 4 provides an overview of study results.

For N staging, the main limitation of conventional and functional imaging relies in the identification of small sized lymph node metastases. A recent study by Heesakkers demonstrated that more than 80% of lymph node metastases presented with size of less than 8mm.(159) Yet, the most promising tracer for N staging remains PSMA. Although other PET tracers such as 11C-Choline, 11C-Acetate or FCH offer some benefits compared to conventional staging, they seem to play only a minor role in light of PSMA-PET/CT.

In the era of new systemic treatment agents, correct identification of distant metastases in newly diagnosed prostate cancer remains crucial and one reason for high failure rates after local treatment might be caused by missed metastases on initial staging. Over the last years, the field of imaging for metastatic disease has rapidly evolved. We found a variety of studies reporting results using radiolabeled PET tracers. Due to heterogeneous study populations, often mixing patients for primary and re-staging purposes as well as for metastatic castration resistant PCa, only a handful of reports met final inclusion criteria for this review.
Overall, modern imaging modalities such as PSMA-, NaF- or FCH-PET/CT as well as WB-MRI have shown superior results compared to conventional imaging; however, direct comparisons of different imaging modalities are missing. Analogous to T and N staging, small sample size and low event rates with wide confidence intervals limit validity of reported results.

The most promising and best-studied tracer represents PSMA. Results from the first RCT demonstrated its superiority to conventional imaging methods. Due to high specificity of the tracer and high tumor-to-background contrast, PSMA makes early identification of bone lesions even before osteolytic or osteoblastic changes possible. (137) PSMA-PET/CT resulted in fewer equivocal results than conventional bone scintigraphy reducing the need for additional testing. (141) By using a single modality such as PSMA-PET/CT, time, radiation dose and costs were spared. A responsible use of resources is essential, not every patient needs whole body work-up. Imaging should be saved for patients at high-risk for metastatic disease while prevalence in low- or early intermediate-risk remains naturally low. (140, 160)

Compared to PET/CT, WB-MRI offers the opportunity of an all-in-one TNM staging without irradiation. However, results for MRI work-up were inconsistent. In addition to previously mentioned limitations by study design, some studies reported excellent performance in detection of bone metastases while in fact other studies observed bone lesions, yet these rarely represented metastases.

There are several limitations of this systematic review. First, due to highly heterogeneous study cohorts, different definition of endpoints, varying imaging techniques, different reader/center experience, and absence of standardized protocols, we had to report our findings in a descriptive manner without pooling of data. Second, the review is limited by the quality of included studies, most being retrospective, lacking direct comparisons to other imaging modalities, including only few patients and having low event rates, resulting in wide confidence intervals and accordingly high level of uncertainty. For T and N staging, histological reference by RP and PLND was required. Histological confirmation of distant metastases was not required; rather, we decided to include studies using at least a best value comparator consisting of imaging, biochemical, and clinical data at baseline and/or follow-up. Finally, although we performed a systematic literature review, some studies might have been missed.
5. Conclusions

A variety of studies on modern imaging techniques for TNM staging in newly diagnosed prostate cancer exist. For T and N staging, reported sensitivity of imaging such as mpMRI or PET/CT varied widely preventing clear recommendations. For M staging, the most promising technique is PSMA-PET/CT. Given the results of our review, most studies were limited by small sample size and low event rate resulting in large confidence intervals and accordingly high level of uncertainty. Therefore, ideally large, prospective studies of 1.) mpMRI and 2.) PET/CT tracers for respectively accurate T-, N- and M-staging using standardized imaging techniques, procedures and appropriate imaging-related reporting systems. Studies would need a clear definition of outcome variables, confirmed by pathological examination and clinical follow-up. For N-staging, predefined templates of anatomical lymph node regions would be necessary to better correlate imaging and pathological results. A central pathological and radiological review with blinding to all data is mandatory. Once, acceptable sensitivity and specificity is achieved, the next step would be an RCT comparing different modalities such as mpMRI and PET/CT to determine the best imaging tool for T and N staging. For M staging, next step would be an RCT comparing PSMA-PET/CT to other modern imaging modalities, especially WB-MRI and/or NaF- and FCH-PET/CT. Furthermore, studies that externally validate the incorporation of mpMRI and/or PET/CT results into existing risk tools are necessary.
References


78. Somford DM, Hamoen EH, Futterer JJ, et al. The predictive value of endorectal 3 Tesla multiparametric magnetic resonance imaging for extraprostatic extension in


91. Zanelli E, Giannarini G, Cereser L, et al. Head-to-head comparison between multiparametric MRI, the partin tables, memorial sloan kettering cancer center nomogram, and CAPRA score in predicting extraprostatic cancer in patients undergoing


Heck MM, Souvatzoglou M, Retz M, et al. Prospective comparison of computed tomography, diffusion-weighted magnetic resonance imaging and 11C choline positron emission tomography/computed tomography for preoperative lymph node staging in


choline positron emission tomography (PET)/computed tomography (CT) and (18) F NaF PET/CT. BJU international. 2014;114(6):818-23.


Figure 1. Flow chart displaying search strategy and study selection following the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) statement guidelines.

Supplementary Figure 1. Risk of bias and study applicability according to QUADAS-2 criteria