Living systematic review and meta-analysis of the prostate MRI diagnostic test with Prostate Imaging Reporting and Data System (PI-RADS) assessment for the detection of prostate cancer: study protocol

Benedict Oerther,1 Christine Schmucker,2 Guido Schwarzer,3 Ivo Schoots,4 August Sigle,5 6 Christian Gratzke,5 Fabian Bamberg,1 Matthias Benndorf

ABSTRACT

Introduction The Prostate Imaging Reporting and Data System (PI-RADS) standardises reporting of prostate MRI for the detection of clinically significant prostate cancer. We provide the protocol of a planned living systematic review and meta-analysis for (1) diagnostic accuracy (sensitivity and specificity), (2) cancer detection rates of assessment categories and (3) inter-reader agreement.

Methods and analysis Retrospective and prospective studies reporting on at least one of the outcomes of interest are included. Each step that requires literature evaluation and data extraction is performed by two independent reviewers. Since PI-RADS is intended as a living document itself, a 12-month update cycle of the systematic review and meta-analysis is planned. This protocol is in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses—Protocols statement. The search strategies including databases, study eligibility criteria, index and reference test definitions, outcome definitions and data analysis processes are detailed. A full list of extracted data items is provided. Summary estimates of sensitivity and specificity (for PI-RADS ≥3 and PI-RADS ≥4 considered positive) are derived with bivariate binomial models. Summary estimates of cancer detection rates are calculated with random intercept logistic regression models for single proportions. Summary estimates of inter-reader agreement are derived with random effects models.

Ethics and dissemination No original patient data are collected, ethical review board approval, therefore, is not necessary. Results are published in peer-reviewed, open-access scientific journals. We make the collected data accessible as supplemental material to guarantee transparency of results.

PROSPERO registration number CRD42022343931.

INTRODUCTION

Prostate MRI has emerged as a fundamental tool in the diagnostic pathway for prostate cancer.1 Recently, it has been strongly recommended by international guidelines for diagnosis in various clinical settings2 3—including biopsy naïve patients and patients with prior negative biopsy and persistent suspicion of prostate cancer. Because of these strong recommendations, the number of prostate MRI examinations performed will substantially increase throughout the next years.

The interpretation of prostate MRI is standardised with a formal lexicon: the Prostate Imaging Reporting and Data System (PI-RADS). PI-RADS was introduced in 2012,4 has been updated to V.2.0 in 20155 and moved to V.2.1 in 2019.6 Analysis of T2-weighted, diffusion-weighted and contrast-enhanced images lead to assessment categories 1 to 5, for single lesions and the entire prostate. The higher the assessment category, the higher the probability of clinically significant cancer. The interpretation lexicon has been updated in each iteration of PI-RADS, meaning
changes in MRI descriptor definition and influence of the
single imaging sequences on final assessment cate-
gories have taken place. The PI-RADS lexicon is explicitly
designed as a living document, meaning that the inter-
pretation lexicon is adapted as evidence about the diag-
nostic performance is generated.

Currently, there is still more evidence regarding the
V.2.0 lexicon as compared with V.2.1 lexicon. Regarding
diagnostic accuracy, in 2017, Woo et al performed a meta-
analyses of 21 studies (3857 patients) using PI-RADS V.2.0
and reported a pooled sensitivity of 89% and a pooled
specificity of 73%.8 For PI-RADS V.2.1, Park et al performed
a similar analysis in 2021 and reported a pooled sensitivity
of 87% and specificity of 74%.9 This initial analysis includes
data from 10 studies and 1240 patients. The cancer detec-
tion rates (CDRs) of PI-RADS V.2.0 have been estimated
with 8% for PI-RADS 2, 13% for PI-RADS 3, 40% for
PI-RADS 4 and 69% for PI-RADS 5.10 For V.2.1 an initial
systematic review and meta-analysis reported CDRs of 2%
for PI-RADS 1, 4% for PI-RADS 2, 20% for PI-RADS 3, 52%
for PI-RADS 4 and 89% for PI-RADS 5 (lesion-level anal-
ysis).11 The PI-RADS lexicon does, in the current edition,
not give numeric definitions of the expected cancer rates in
the assessment categories. Furthermore, no management
recommendations are linked to the assessment categories.

To account for the continuously generated evidence of the
diagnostic performance of PI-RADS and expected
future iterations of the lexicon (with changes in descriptor
definitions and assessment category definitions, and,
therefore, expected changes in diagnostic performance),
we want to establish a living systematic review and meta-
analyses. This living review will estimate the diagnostic
accuracy of the current PI-RADS (sensitivity and speci-
ficity), the cancer detection rates (CDRs) of the assess-
ment categories and inter-reader agreement of category
assignment. We plan to perform update searches and
analyses in 12-month cycles.

Our objective is the implementation of a living
systematic review and meta-analysis of the diagnostic
performance of prostate MRI with PI-RADS assessment
(intervention, V.2.1 and upcoming versions considered)
for the detection of prostate cancer (outcome) in patients
with suspicion for prostate cancer (participants). Diag-
nostic performance of prostate MRI will not be compared
with another diagnostic test (comparator), reference
standard is histopathology.

Methods and analysis

Study design and registration

This is a systematic review protocol, it follows the
Preferred Reporting Items for Systematic Reviews and
Meta-Analyses—protocols (PRISMA-P) guidelines and
format.12 The systematic review has been registered in
PROSPERO. The PRISMA-P checklist for our protocol is
enclosed as an online supplemental file.

Study eligibility criteria

We include prospective and retrospective studies
reporting on the diagnostic accuracy, and/or cancer
detection rates of PI-RADS and/or inter-reader agree-
ment of PI-RADS rating, starting with PI-RADS V.2.1.
Studies that use older versions of the lexicon are not
considered. Studies reporting on a subset of PI-RADS
categories are eligible. We consider studies published as
full text in English. Date restriction is applied, considered
studies need to be conducted in 2019 or later, that is after
the release of the current PI-RADS V.2.1. Studies are still
considered as eligible if included patients were examined
prior to this date but have been reinterpreted by blinded
readers according to the current PI-RADS.

Study population

Our target populations are men with suspicion for pros-
tate cancer, either biopsy naïve or with a prior negative
biopsy. Biopsy naïve patients have a higher pretest prob-
ability for clinically significant cancer.13 Biopsy status
will be considered as a covariate in our analysis. Patients
with known malignancy at the date of prostate MRI or
with prior treatment of the prostate are not considered
eligible.

Index test

Prostate MRI read according to the current PI-RADS
(V.2.1 at the time of writing this protocol) is the diag-
nostic test of interest. We record MRI parameters of
single studies to account for deviations from the proposed
imaging protocol.14 Experience of the involved radiolo-
gist(s) is recorded. We document whether MRI reading
is performed without knowledge of the histopathological
result. We investigate diagnostic performance on lesion
level (up to four lesions per patient are possible) and
patient level (equals highest assigned lesion category
compared with overall histopathological result).

Comparators

Diagnostic accuracy and cancer detection rates of
PI-RADS will not be compared with another diagnostic
test.

Reference test

Histopathological verification of suspicious lesions and
the prostate can be performed in several ways. The type
of targeted lesion biopsy is recorded (cognitive fusion,
transrectal ultrasound MRI fusion, transperineal MRI
ultrasound fusion, in-bore). A systematic biopsy and
additional MRI-directed perilesional biopsies may also
be performed. We record the type and result of targeted
biopsy, type of systematic biopsy (if any) and type of
perilesional biopsies (if any). Histopathological upgrade
of targeted biopsies given the information from system-
atic biopsy is recorded. Furthermore, analysis of pros-
tectomy specimen is eligible as reference standard.

Outcomes

Primary outcome is the detection (sensitivity and spec-
ificity, cancer detection rates) of clinically significant
cancer. The most widely adapted procedure in the litera-
ture regarding PI-RADS is to consider any occurrence of a
histopathological Gleason pattern $\geq 3+4$ as clinically significant.\textsuperscript{10,11} The PI-RADS lexicon offers a more elaborate definition, which is more challenging to establish in clinical routine: 'Gleason score $\geq 7$, including 3+4 with prominent but not predominant Gleason 4 component, and/or volume $>0.5$ cc and/or extraprostatic extension'.\textsuperscript{14} Especially the last point is, given that histopathological verification is performed by targeted lesion biopsies, systematic biopsies (this is the case in the majority of individual cases and studies), often not possible to establish prior to surgery. Type of definition of clinically significant cancer will be considered as a covariate. Analysis is performed on lesion level (each lesion observed in the MRI examination, up to four lesions per patient, targeted biopsy as reference standard; studies reporting only the results of targeted biopsies without additional systematic biopsies are eligible for the lesion-level analysis only) and patient level (highest PI-RADS category as index test, lesion and systematic biopsy and (if performed) perilesional biopsy or prostatectomy as reference standard).

Secondary outcomes are the detection (sensitivity and specificity, cancer detection rates) of insignificant cancer, any cancer, Gleason $\geq 4+3$ (if reported) and $\geq 5+4$ with cribriform growth pattern (if reported). Although the PI-RADS lexicon explicitly does not aim at the detection of clinically insignificant cancer, knowledge about occurrence of these cancers is still important from a public health perspective. Patients with a diagnosis of clinically insignificant cancer will be closely monitored with active surveillance, including serial prostate-specific antigen (PSA) testing, MRI and biopsies.\textsuperscript{15} For primary outcome and secondary outcomes, we investigate the scenarios PI-RADS $\geq 3$ and $\geq 4$ considered positive for the estimation of sensitivity and specificity.

Inter-reader agreement of lesion and patient classification with PI-RADS (Cohen’s kappa values) is defined as a secondary outcome.

**Information sources and search strategy**

We search the following databases for published studies, ongoing studies or completed studies not (yet) published: MEDLINE, Embase, Cochrane Library, ISRCTN, ClinicalTrials.gov, ICTRP and Deutsches Register Klinischer Studien (DRKS). Time restriction will be applied. We consider all studies conducted from March 2019 onwards—PI-RADS V.2.1 has been published in March 2019. Bibliographies of included articles will be manually checked for further eligible studies. The search strategy will be reused for the planned update cycles in the living systematic review framework.

Our MEDLINE search is structured as follows:

\((\text{PI-RADS} \text{ OR } (\text{PI-RADS}) \text{ OR } (\text{"prostate imaging reporting and data system"})) \text{ AND } (\text{"2019/03/01" [Date - Publication]: "3000/12/12" [Date - Publication]}))\). Searches of the other databases are adapted accordingly. Full search strategies of all databases are provided as an online supplemental file to this protocol.

**Data management**

Search results from the different databases are combined in a dedicated software environment (eg, Rayyan, https://www.rayyan.ai/), duplicates will be removed. Backup copies are generated after the single database searches.

**Selection process**

Two independent reviewers evaluate eligibility of search results. First, selection is performed on title and abstract basis. Studies considered relevant (or potentially relevant) based on title and abstract screening are further considered based on their full text (full-text screening). In each step, discrepancies will be resolved by discussion and by consultation of a third reviewer, if needed. The reason for exclusion is recorded in each selection step.

**Data collection process**

Two independent reviewers extract data from the included studies in duplicate spreadsheets with predefined data items. We define a core set of data items (compare for tables 1 and 2). If any items of this set are missing, authors of primary studies are contacted (at least two times) to obtain this missing data.

**Risk of bias assessment**

For the evaluation of risk of bias and applicability of results (study-level analysis each) the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) framework is used.\textsuperscript{16} Two independent reviewers evaluate risk of bias and applicability of results in the domains patient selection, index test, reference standard and flow and timing (the latter not for applicability evaluation). Discrepancies are resolved by discussion and by consultation of a third reviewer, if needed. From the results of the QUADAS-2 analysis, we will infer the overall risk of bias for obtained results. Studies are not excluded from data synthesis based on the QUADAS-2 evaluation alone.

**Data synthesis and statistical analysis**

Data describing patient populations of the included studies (eg, mean age, mean PSA value, mean prostate volume, prior biopsy status) are presented in table format. Data synthesis of outcomes (diagnostic accuracy in terms of sensitivity and specificity, cancer detection rates, inter-reader agreement in terms of Cohen’s kappa values) is performed given a set of homogeneous studies is identified. The required minimum set of homogeneous study characteristics is: (1) reading of prostate MRI is performed without knowledge of the histopathological results, (2) MRI is performed according to PI-RADS recommendations, (3) for inter-reader agreement, comparable metrics are reported.

We derive pooled estimates of sensitivity and specificity with bivariate binomial models.\textsuperscript{17} A summary receiver operating characteristic (ROC) curve with a 95% confidence region is derived for graphical representation. We examine the scenarios with PI-RADS $\geq 3$ and PI-RADS $\geq 4$ considered as a positive test on lesion level and patient level (overall, four scenarios). Possible publication bias is
visually assessed with funnel plots, Deek’s test will be used to test for asymmetry.\textsuperscript{18} Coupled forest plots of sensitivity and specificity and correlation between sensitivity and 1-specificity are analysed for assessment of heterogeneity of results.\textsuperscript{19}

We expect cancer rates in the assessment categories to vary across studies; partly because of different local reading standards, partly because of local differences/thresholds for referral to prostate MRI and targeted biopsy, partly because of different pretest probabilities and, thus, differences in the patient cohorts examined. In other words, we assume a certain degree of clinical and methodological heterogeneity between studies and do not expect results to vary because of random sampling error alone. For this reason, we employ random intercept logistic regression models for meta-analysis of single proportions to derive summary estimates for cancer detection rates of the PI-RADS categories\textsuperscript{20} and subcategories of PI-RADS 3 and 4. Heterogeneity of reported cancer detection rates is assessed with Higgins’ \textit{I}^2 statistic, with \textit{I}^2>50\% denoting substantial heterogeneity.\textsuperscript{19}

Meta-regression with the following covariates (if data are sufficient) is performed to examine possible causes of heterogeneity (diagnostic accuracy and cancer detection rates): type of study population (prior biopsy status), magnetic field strength, multiparametric versus biparametric MRI, definition of clinically significant cancer, type of lesion verification, lesion localisation (peripheral zone vs transition zone), reader experience, pretest...
probability and mean/median PSA in the study population. Subgroup analyses of covariates are performed for univariate analyses.

The summary measure for inter-reader agreement (Cohen’s kappa values) will be derived with a random effects model. This follows the method proposed by Sun.21 We examine the role of reader experience as a covariate—two highly experienced readers can be expected to agree more often compared with two relatively inexperienced readers or two readers with different levels of experience.

If quantitative data synthesis is not considered appropriate for one or more defined outcomes, a synopsis of findings is given in table format. Order of presentation is stratified by risk of bias and definition of clinically significant cancer used.

All statistical analyses are conducted using R (https://www.R-project.org/).22

**GRADE assessment**
Quality of evidence per outcome is analysed according to the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) System,23 results from the QUADAS-2 analysis are used for risk of bias assessment in this context. Certainty of evidence is rated as high, moderate, low or very low. Results are made available in a summary of findings table.

**Patient and public involvement**
In the development phase of the project, the Bundesverband Prostatakrebs Selbsthilfe e.V. (https://prostatakrebs-bps.de/) was involved in defining relevant

---

**Table 2** Extracted data items—outcome data

<table>
<thead>
<tr>
<th>Data item(s)</th>
<th>Levels</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcome data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Definition of csCA used</td>
<td>PI-RADS Lexicon definition, other definition</td>
<td>Exact definition is recorded</td>
</tr>
<tr>
<td>Number of lesions and patients with csCA*</td>
<td>Numerical</td>
<td></td>
</tr>
<tr>
<td>Number of lesions and patients with ncsCA</td>
<td>Numerical</td>
<td></td>
</tr>
<tr>
<td>Number of lesions and patients with Gleason score≥4+3</td>
<td>Numerical</td>
<td></td>
</tr>
<tr>
<td>Number of lesions and patients with Gleason score≥3+4 and cribriform growth pattern</td>
<td>Numerical</td>
<td></td>
</tr>
</tbody>
</table>
| Sensitivity and specificity* | Numerical | ▶ Reported in paper or reconstructed from presented data, 2×2 contingency tables from paper or reconstructed are recorded  
▶ Scenarios PI-RADS≥3 and PI-RADS≥4 considered positive are examined  
▶ Data are extracted on lesion level and patient level, for all extracted definitions of prostate cancer (csCA, ncsCA, Gleason score≥4+3 and ≥ 3+4 with cribriform growth pattern) |
| Cancer detection rates* | Numerical | ▶ Number of malignant cases in each reported PI-RADS category divided by all cases in each PI-RADS category  
▶ Data are extracted on lesion level and patient level, for all extracted definitions of prostate cancer (csCA, ncsCA, Gleason score≥4+3 and ≥ 3+4 with cribriform growth pattern)  
▶ Subcategories in PI-RADS 3 and 4 are recorded separately, if information is available  
▶ For low PI-RADS categories, the information will also be expressed as negative predictive value |
| Reader agreement* | Type of obtained inter-reader agreement metric, numerical value of metric |

*Core data items (reporting of at least one defined outcome is required). If missing, authors of the primary studies are contacted two times to obtain the missing data.

csCA, clinically significant cancer; ncsCA, clinically non-significant/insignificant cancer; PI-RADS, Prostate Imaging Reporting and Data System.
research questions. The Bundesverband Prostatakrebs Selbsthilfe e.V. agreed to disseminate results in their network of support groups.

**Living review framework**

We plan to implement a 12-month cycle to update our literature search, study selection and data analysis. This is because an accumulation of evidence about the diagnostic performance of PI-RADS can be expected, especially for subcategories in categories 3 and 4. Furthermore, PI-RADS is itself intended as a living diagnostic algorithm—that is, new iterations can be expected. Given that the diagnostic algorithm is further adapted, changes in diagnostic accuracy can be expected. If a new version of PI-RADS is released, our literature search strategy will remain unchanged. Data collection and reporting of results will pertain to the current version of PI-RADS.

We consider the living systematic review framework suitable for our project because the scope and needs address the three demands as expressed in the initial discussion of living systematic reviews by Elliott et al.1: 1. up-to-date information is important for decision-making; for informed, shared decision-making how to proceed with the result of a prostate MRI examination, accurate estimates of diagnostic accuracy of PI-RADS and cancer detection rates of the categories are crucial. Furthermore, management recommendations are planned to be linked to assessment categories in future versions of PI-RADS. Before recommending biopsy, for example, there needs to be an established expected cancer rate for a certain assessment category.

2. Certainty in the existing evidence is low: at the moment, we have limited evidence (meta-analyses do exist for diagnostic accuracy and cancer detection rates of PI-RADS V.2.1; however, they include relatively few patients). Furthermore, we see a need to systematically review the performance of subcategories in PI-RADS categories 3 and 4.

3. There will be new research evidence: the publication field of prostate MRI and PI-RADS is highly dynamic, the number of relevant papers is increasing at a fast rate. We expect new accumulating evidence especially for subcategories (different lesion entities in categories 3 and 4). Furthermore, new evidence will be generated given a new iteration of PI-RADS is published. A timely evidence synthesis is warranted in this case.

Our search strategy and data used for analyses will be published as online supplemental file to the systematic review and meta-analysis.

**Ethics and dissemination**

No original data are collected in this systematic review and meta-analysis, ethical review board approval, therefore, is not required. Results are published in peer-reviewed, open-access scientific journals. We make the collected data accessible as online supplemental materials to guarantee transparency of results.

**DISCUSSION**

With the recently put forward strong recommendations for prostate MRI prior to biopsy in various national and international guidelines, a rapidly increasing volume of prostate MRI examinations can be expected in the next years. The increasing number of examinations performed requires a standardised, evidence-based diagnostic workflow to streamline patient management.

PI-RADS, having been established in 2012, offers this standardisation. PI-RADS provides a universally understood reporting language on the descriptor level and works well as a risk stratification tool for clinically significant prostate cancer. For V.2.0, a systematic review and meta-analysis of inter-reader agreement reported an overall moderate to substantial agreement for PI-RADS category assignment. The diagnostic accuracy of PI-RADS has been subjected to a multitude of studies—initial estimates for sensitivity, specificity and the cancer detection rates are available for V.2.1. Before recommending biopsy, for example, there needs to be an established expected cancer rate for a certain assessment category.

As evidence about the diagnostic performance of PI-RADS accrues, these estimates will become more precise. Or, given considerable heterogeneity of estimates between studies, the identification of covariates that affect diagnostic accuracy and cancer detection rates becomes possible. This knowledge could ultimately be included into PI-RADS itself or future guidelines.

At the moment, assessment categories 3 and 4 are assigned to a heterogeneous group of lesions each. For example, in the transition zone assessment category 3 comprises lesions with different appearance in T2-weighted images (atypical nodules and heterogenous lesions with obscured margins). Costa et al report a cancer rate of 6% and 11% for these two lesion types, although this difference is not statistically significant in their study. If there are systematic differences of cancer rate between lesion subtypes in the same PI-RADS assessment category, this might influence the planned linking of management recommendations to assessment categories.

Our living systematic review framework establishes an evidence base for precise estimates of diagnostic accuracy of the current PI-RADS (with different thresholds considered positive), the cancer detection rates of assessment categories and subcategories and inter-reader agreement. The results can be employed by urologists, radiologists and patients for decision-making after prostate MRI and help in the development of PI-RADS itself and future guidelines.

**Author affiliations**

Department of Diagnostic and Interventional Radiology, Medical Center-University of Freiburg, Freiburg, Germany
REFERENCES