Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Diagnostic Accuracy of Prostate-Specific Membrane Antigen PET/CT for Pelvic Lymph Node Staging in Intermediate-to-High-Risk Prostate Cancer: A Systematic Review and Meta-Analysis

Bahareh Mehdikhani
Iran University of Medical Sciences
Reza Shahrokhi Damavand
subgene87@gmail.com

Guilan University of Medical Sciences
Ehsan Zolfi
Iran University of Medical Sciences
Nima Narimani
Iran University of Medical Sciences
Mohammad Mehdi Atarod
Iran University of Medical Sciences
Hadi Maleki
Alborz University of Medical Sciences
Saman Farshid
Urmia University of Medical Sciences
Alireza Negahi
Iran University of Medical Sciences
Kazem Aghili
Shahid Sadoughi University of Medical Sciences and Health Services
Hossein Neamatzadeh
Shahid Sadoughi University of Medical Sciences and Health Services

Research Article

Keywords: PSMA PET/CT, prostate cancer, lymph node staging, N1 disease, diagnostic accuracy, meta-analysis, extended pelvic lymph node dissection, 68Ga-PSMA-11, bivariate model

Posted Date: July 31st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10022523/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background.

Accurate pelvic lymph node (N1) staging governs treatment selection in intermediate-to-high-risk prostate cancer, yet conventional imaging is insensitive to small nodal deposits. Prostate-specific membrane antigen positron emission tomography/computed tomography (PSMA PET/CT) is increasingly used for primary staging, but its per-patient accuracy against a histopathologic reference standard—and the factors that modulate it—remain incompletely defined.

Methods.

Following PRISMA-DTA 2020 guidance and a prospectively registered protocol, we searched five bibliographic databases from inception to June 8, 2026, with registry, grey-literature, and citation searches. Eligible studies evaluated PSMA PET/CT for primary per-patient N1 staging in treatment-naïve intermediate-to-high-risk disease against histopathologic extended pelvic lymph node dissection (ePLND). We fitted bivariate random-effects models, derived hierarchical and Moses–Shapiro–Littenberg summary ROC (SROC) curves, and performed prespecified subgroup analyses, meta-regression, 14 sensitivity analyses, leave-one-out and cumulative meta-analyses, and publication-bias assessment.

Results.

Thirty studies (5,927 patients; nodal prevalence 25.5%) were analyzed. Pooled sensitivity was 56.9% (95% CI 48.3–65.0%; I^2 = 86.9%) and specificity 93.2% (92.1–94.2%; I^2 = 26.0%), with a diagnostic odds ratio of 19.03 (12.82–28.23), positive and negative likelihood ratios of 7.44 and 0.50, and SROC area under the curve 0.90 (Q^* = 0.895). The sensitivity prediction interval was wide (20.2–87.2%) but that for specificity narrow (89.8–95.6%). 68Ga-PSMA-11 achieved the highest sensitivity (65.1%) and 18F-DCFPyL the lowest (34.0%), while specificity was uniform across tracers. Partial verification inflated apparent sensitivity (81.7% vs 49.4% with complete verification). Meta-regression identified median PSA (p = 0.017), partial verification (p = 0.007), and prior conventional imaging (p = 0.006) as independent predictors of sensitivity. Estimates were robust across all 14 sensitivity analyses (sensitivity 43.1–65.1%; specificity 92.9–94.0%); significant small-study effects were present (Deeks' p = 0.0009).

Conclusions.

PSMA PET/CT couples high, reproducible specificity with moderate, heterogeneous sensitivity for pelvic N1 staging. A positive scan reliably confirms nodal disease and supports treatment intensification, whereas a negative scan cannot exclude micrometastatic involvement and should be integrated with validated nomograms and selective ePLND rather than used in isolation to guide surgical decision-making.

1. Introduction

Prostate cancer is the most frequently diagnosed malignancy and a leading cause of cancer death among men in Western countries; in the United States alone, an estimated 313,780 new cases and 35,770 deaths were projected for 2025, accounting for roughly 30% of male cancers [1]. Importantly, after nearly a decade of decline, incidence has reversed to an annual rise of approximately 3%—driven disproportionately by advanced-stage disease, which is increasing by 4.6–4.8% per year—so a growing proportion of men now present with tumors in which accurate locoregional staging directly governs management [2]. Although organ-confined disease carries an excellent prognosis after curative-intent therapy, pelvic lymph node metastasis (N1, pN1) marks a decisive prognostic threshold that reshapes treatment—prompting the addition of androgen-deprivation therapy to radiotherapy, extension of radiation fields to the regional lymphatics, or modification of the surgical lymphadenectomy template [3, 4]. Reliable preoperative nodal staging is therefore central to contemporary risk-adapted care, and the extended pelvic lymph node dissection (ePLND) that establishes pathologic nodal status remains both the diagnostic reference standard and a procedure of debated therapeutic value that carries a non-trivial burden of lymphocele, lymphedema, and thromboembolic complications [5].

Conventional cross-sectional imaging performs poorly in this role: pooled estimates indicate that computed tomography and magnetic resonance imaging achieve nodal sensitivities below 40% because they depend on nodal size and morphology, missing the small-volume deposits that dominate newly diagnosed disease [6]. PSMA-targeted PET imaging has transformed this landscape. PSMA is a type II transmembrane glycoprotein markedly overexpressed by prostate adenocarcinoma, with expression rising in step with tumor grade, androgen independence, and metastatic potential [7]. Several radioligands are in clinical use—the gallium-68–labeled agents 68Ga-PSMA-11 and 68Ga-PSMA-I&T and the fluorine-18–labeled agents 18F-DCFPyL, 18F-PSMA-1007, and 18F-rhPSMA-7—each with distinct pharmacokinetics, urinary-excretion profiles, and logistic characteristics [8].

The randomized proPSMA trial established the superiority of PSMA PET/CT over conventional imaging for detecting nodal or distant metastasis in high-risk disease, reporting 27% greater accuracy and a change in management in 28% of patients [9]; these findings, reinforced by the OSPREY trial of 18F-DCFPyL [10] and by the maturation of PSMA-directed radioligand therapy into a diagnostic–therapeutic continuum [11], underpinned rapid regulatory and guideline incorporation. The 2024 EAU and NCCN v1.2025 guidelines now endorse PSMA PET/CT for metastatic screening in high-risk disease and as an option in unfavorable intermediate-risk disease, with the AUA/SUO and SNMMI frameworks broadly concordant [3, 12–14]. A recurring caveat across these documents is that nuclear-imaging findings—false-positives in particular—warrant histologic confirmation whenever feasible, reflecting the consequences of staging error in surgical candidates.

A fundamental limitation tempers this enthusiasm: PSMA PET/CT detectability is governed by metastatic deposit size. Multiple histology-correlated cohorts report that most pathologically positive nodes smaller than 5 mm escape detection, with sensitivity rising steeply once deposits exceed that threshold [15, 16]. Whether—and how—a negative PSMA PET/CT should influence the decision to perform or omit ePLND therefore remains contested, and emerging strategies including PET-guided lymphadenectomy [17], artificial-intelligence-assisted lesion detection [18], and combined PSMA PET/MRI [19] all depend on an accurate baseline estimate of per-patient sensitivity against histopathology. Prior syntheses have not resolved this

question: several pool per-lesion and per-patient data, accept composite or imaging-based reference standards, or combine treatment-naïve and biochemical-recurrence populations, each of which biases accuracy estimates for the specific scenario of preoperative N1 staging [20]. The recent comprehensive review by Mazzone and colleagues, whose lymph-node sub-analysis of dissection-referenced cohorts reported pooled sensitivity of 54% and specificity of 94%, nonetheless spanned heterogeneous populations and analysis levels [21].

We therefore conducted a systematic review and meta-analysis with deliberately stringent eligibility criteria—restricting inclusion to per-patient analyses with histopathologic ePLND reference standards in treatment-naïve intermediate-to-high-risk disease—to provide a low-bias estimate of the diagnostic accuracy of PSMA PET/CT for pelvic N1 staging. Secondary objectives were to characterize the sources of heterogeneity through subgroup and meta-regression analyses, to quantify the influence of partial-verification bias and tracer choice, to test the robustness of pooled estimates across sensitivity analyses, and to evaluate publication bias.

2. Materials and Methods

This review followed the PRISMA 2020 statement, its diagnostic-test-accuracy extension (PRISMA-DTA), and the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy [22–24]. The protocol was prospectively registered with PROSPERO before screening began, and no substantive deviations from the registered plan occurred. The patient was the unit of analysis, and the target condition was histopathologically confirmed pelvic lymph node metastasis (pN1).

2.1 Search Strategy

A senior medical librarian designed and executed the search across five bibliographic databases—MEDLINE (via PubMed), Embase (via Elsevier), the Cochrane Central Register of Controlled Trials (CENTRAL), Web of Science Core Collection, and Scopus—from inception to June 8, 2026. These were supplemented by structured searches of ClinicalTrials.gov and the WHO International Clinical Trials Registry Platform for completed trials with posted results, Google Scholar for grey literature, and backward and forward citation chasing of all included studies and relevant prior reviews. No date or language restrictions were applied at the search stage; non-English records without an available translation were excluded only at full-text screening and enumerated.

The strategy combined controlled vocabulary (MeSH and Emtree) with free-text terms across four concept blocks joined by AND, with synonyms within each block joined by OR. Block 1 (index test) comprised “prostate-specific membrane antigen,” PSMA, the individual radioligands, and PET/CT terms; Block 2 (condition) prostate cancer and its synonyms; Block 3 (target) lymph node, nodal, N-staging, and lymphadenectomy terms; and Block 4 (reference/outcome) diagnostic accuracy, sensitivity, specificity, and ePLND/histopathology terms. The strategy was peer-reviewed against the PRESS (Peer Review of Electronic Search Strategies) checklist before execution, and all retrieved records were imported into reference-management software for automated and manual de-duplication. The complete line-by-line strategy for each database is available from the corresponding author.

2.2 Eligibility Criteria

2.2.1 Inclusion criteria

Studies were eligible if they satisfied all of the following: (i) a diagnostic-accuracy, prospective cohort, retrospective cohort, or randomized design evaluating any PSMA-targeting radiotracer (68Ga-PSMA-11, 68Ga-PSMA-I&T, 68Ga-PSMA-HBED-CC, 18F-PSMA-1007, 18F-DCFPyL, or 18F-rhPSMA-7); (ii) a population of treatment-naïve patients with newly diagnosed intermediate- or high-risk prostate cancer defined by D'Amico, EAU, NCCN, or ISUP grade group 2–5 criteria; (iii) PSMA PET/CT performed for primary staging before curative-intent prostatectomy or radiotherapy; (iv) histopathologic examination of ePLND specimens (minimum template encompassing the external iliac, internal iliac, and obturator regions) as the reference standard; (v) sufficient data to construct a per-patient 2×2 contingency table, or sensitivity and specificity reported with extractable patient denominators; (vi) pelvic lymph node metastasis (pN1) as the target condition; and (vii) per-patient analysis. Where a 2×2 table was not published directly, it was algebraically reconstructed from reported sensitivity, specificity, prevalence, and total N, cross-validated for internal consistency, flagged, and examined in sensitivity analysis.

2.2.2 Exclusion criteria

Studies were excluded for any of: evaluation in the biochemical-recurrence or castration-resistant setting; a reference standard other than histopathologic ePLND (including composite or imaging-only standards); per-lesion or per-region analysis without patient-level data; a mixed primary-staging and recurrence cohort without separable data; a pN1-only or pN0-only series precluding calculation of both sensitivity and specificity; a tracer-comparison or technical report lacking N1 accuracy against ePLND; distant metastasis (M1, including M1b) as the target condition; insufficient data to reconstruct a 2×2 table; reviews, editorials, letters, case reports, or conference abstracts without original patient-level data; and fewer than 20 evaluable patients. Full-text exclusions and their reasons are catalogued in Supplementary Table S1.

2.3 Study Selection and Data Extraction

Two reviewers independently screened titles and abstracts and then full texts against the eligibility criteria, resolving disagreements by discussion or adjudication by a third reviewer; inter-reviewer agreement at the full-text stage was quantified with the Cohen κ statistic. A piloted, standardized extraction form—tested on a random sample of five studies before full deployment—captured study, patient, index-test, and reference-standard characteristics, the per-patient 2×2 table, and reported accuracy metrics. Two reviewers independently verified every 2×2 table against the source publication; calculated and reported metrics were reconciled, and discrepancies exceeding two percentage points triggered re-extraction and, when necessary, direct contact with study authors. Internal consistency was confirmed by recomputing sensitivity, specificity, predictive values, and accuracy from the extracted cell counts, and any cell count of zero was retained for the sensitivity-analysis dataset and handled with a continuity correction during pooling (Section 2.5).

2.4 Quality Assessment

Two reviewers independently appraised methodological quality with QUADAS-2 across the patient-selection, index-test, reference-standard, and flow-and-timing domains, rating each as low, high, or unclear risk of bias

and additionally judging the applicability concerns specified by the tool [25]. Studies with any domain rated high or unclear were carried forward into sensitivity analyses, and studies meeting all four domains at low risk were designated high quality. The certainty of the body of evidence for the primary outcomes was further considered using a GRADE-informed framework for diagnostic accuracy. Domain-level and study-level judgments are reported in the Results and tabulated in Supplementary Table S2.

2.5 Statistical Analysis

Analyses were performed in R 4.4.0 (packages mada, meta, and lme4) and Python 3.12 (NumPy, SciPy, statsmodels). The primary synthesis used the bivariate random-effects model of Reitsma and colleagues, which jointly models logit-sensitivity and logit-specificity while accounting for their negative correlation and for between-study heterogeneity, estimated by restricted maximum likelihood [26]; the mathematically equivalent hierarchical SROC formulation of Rutter and Gatsonis was used to derive the summary curve [27]. For parameters without a closed bivariate form (predictive values, likelihood ratios, and the diagnostic odds ratio), DerSimonian–Laird random-effects pooling was applied on the appropriate transformed scale with a 0.5 continuity correction for zero cells, and estimates were back-transformed to the natural scale [28]. The SROC curve was additionally characterized with the Moses–Shapiro–Littenberg model, reporting the Q* index (the point of equal sensitivity and specificity) and the area under the curve [29].

Heterogeneity was quantified with Cochran's Q and the I^2 statistic—interpreted at 25%, 50%, and 75% as low, moderate, and high—together with the between-study variance τ^2 and 95% prediction intervals derived from the bivariate covariance structure to convey the plausible range of accuracy in a future study [30]. Prespecified subgroup analyses examined tracer type, study design, center type, sample-size stratum, verification status, methodological-quality tier, data-source type (validated versus reconstructed 2×2), image-analysis method, reader blinding, geographic region, publication period, risk-group definition, reference-standard type, and disease-prevalence quartile; subgroups containing a single study were reported descriptively and not pooled. Univariable meta-regression related continuous covariates (log sample size, publication year, disease prevalence, median PSA, median age) and dichotomous covariates (design, verification, quality, data source, tracer class, prior conventional imaging, composite reference) to logit-sensitivity, logit-specificity, and log-DOR using weighted least squares, with R^2 quantifying the proportion of between-study variance explained by each covariate.

Robustness was assessed through 14 prespecified sensitivity analyses—excluding, in turn, reconstructed 2×2 tables, small studies ($N < 50$), partially verified studies, and lower-quality studies; restricting to prospective, retrospective, 68Ga-PSMA-11, or 18F-based studies; restricting to completely verified plus directly validated data; and omitting individual influential studies—complemented by leave-one-out diagnostics and by cumulative meta-analysis ordered by publication year. Publication bias and small-study effects were evaluated with Deeks' funnel-plot asymmetry test (the recommended test for diagnostic meta-analyses), Egger's regression, and Begg's rank correlation, with trim-and-fill estimation of potentially missing studies [31–33]. Two-sided $p < 0.05$ denoted significance, except for Deeks' test, where $p < 0.10$ indicated asymmetry.

3. Results

3.1 Study Selection

The search returned 2,290 records (2,245 from bibliographic databases and 45 from registry, grey-literature, and citation sources). After removal of 610 duplicates, 1,680 records were screened by title and abstract and 1,580 were excluded. One hundred full texts were assessed for eligibility, of which 16 were excluded with documented reasons (Supplementary Table S1). Forty studies entered qualitative synthesis; 30 provided complete per-patient 2×2 data and constituted the primary bivariate meta-analysis (5,927 patients, 2016–2026), while 10 contributed to subgroup or sensitivity analyses only (Supplementary Table S6). The selection process is summarized in the PRISMA-DTA 2020 flow diagram (Fig. 1).

3.2 Study and Patient Characteristics

The 30 primary studies spanned 14 countries and were predominantly single-center (24/30) and retrospective (20/30); 10 were prospective and 6 were multi-center. Sample sizes ranged from 30 to 873 patients (median 99), and nodal prevalence ranged from 10.2% to 42.3% (pooled 25.5%). 68Ga-PSMA-11 was the most frequent tracer (15 studies), followed by other 68Ga formulations (6), 18F-PSMA-1007 (5), 18F-DCFPyL (3, including the OSPREY [10] and SALT [45] cohorts), 18F-rhPSMA-7 (1), and mixed-tracer cohorts (3). All studies used pelvic lymph node dissection—ePLND in 28 and standard PLND in 2—as the histopathologic reference standard; 25 verified every enrolled patient and 5 employed partial verification. Median PSA at staging, where reported, ranged from approximately 8 to 24 ng/mL across cohorts. Full study characteristics with reference citations are provided in Table 1, and the per-study 2×2 contingency data with derived accuracy metrics in Table 2.

Table 1
Characteristics of the Included Studies.

Study	Country	Design	Centers	N	Verified	Risk Group	Tracer
Budäus 2016 [34]	Germany	Retro, SC	1	30	30	High	68Ga-PSMA
Maurer 2016 [35]	Germany	Retro, SC	1	130	130	High	68Ga-PSMA
van Leeuwen 2016 [36]	Netherlands	Retro, SC	1	30	30	High	68Ga-PSMA
Obek 2017 [37]	Turkey	Retro, MC	Multi	51	51	High/VH	68Ga-PSMA
Zhang 2017 [38]	China	Retro, SC	1	42	42	Int–High	68Ga-PSMA-11
Yaxley 2019 [39]	Australia	Retro, SC	1	208	208	Int/High	68Ga-PSMA
Kroenke 2020 [40]	Germany	Retro, SC	1	58	58	High	18F-rhPSMA-7
Esen 2021 [41]	Turkey/Germany	Retro, SC	1	96	96	Int/High	68Ga-PSMA-11
Hope 2021 [42]	USA	Prosp, MC	Multi	277	277	High	68Ga-PSMA-11
Malaspina 2021 [43]	Italy	Prosp, SC	1	79	79	Int/High	18F-PSMA-1007
Meijer 2021 [44]	International	Retro, MC	3	757	757	Mixed	Mixed
Pientka 2021a [10]	USA	Prosp, MC	Multi	252	252	High	18F-DCFPyL
Pientka 2021b [45]	USA	Prosp, MC	Multi	117	117	Int/High	18F-DCFPyL
Szigeti 2021 [46]	Austria	Prosp, SC	1	81	48	Int/High	68Ga-PSMA-11
van Kalmthout 2021 [47]	Netherlands	Prosp, SC	1	103	97	High	68Ga-PSMA-11
Erdem 2022 [48]	Turkey	Retro, SC	1	49	49	Int/High	68Ga-PSMA-11

Retro, retrospective; Prosp, prospective; SC, single-center; MC, multi-center; VH, very-high-risk; Int, intermediate; ePLND, extended pelvic lymph node dissection; PLND, pelvic lymph node dissection; 18F-1007, 18F-PSMA-1007.

Study	Country	Design	Centers	N	Verified	Risk Group	Tracer
Hermsen 2022 [49]	Netherlands	Prosp, SC	1	99	99	Int–High	18F-PSMA-1007
Jakobsen 2022 [50]	Sweden	Retro, SC	1	104	104	Int/High	18F-PSMA-1007
Kopp 2022 [51]	Germany	Retro, SC	1	39	39	Int	68Ga-PSMA-11
Maestroni 2023 [52]	Italy	Retro, SC	1	42	42	Int/High	68Ga-PSMA-11
Di Giorgio 2024 [53]	Italy	Retro, SC	1	78	78	Int/High	68Ga-PSMA
Ozturk 2024 [54]	Turkey	Retro, SC	1	583	583	Int/High	68Ga + 18F-DCFPyL
Patel 2024 [55]	India	Prosp, SC	1	30	30	High	68Ga-PSMA-11
Rosales 2024 [56]	Spain	Retro, SC	1	122	89	Int/High	68Ga-PSMA-11
Rovera 2024 [57]	Italy	Prosp, SC	1	60	32	High	68Ga-PSMA-11
Arena 2025 [58]	Italy	Retro-prosp, SC	1	156	156	Int/High	68Ga + 18F-1007
Esen 2025 [59]	Turkey/Germany	Retro, SC	1	521	521	Int/High	68Ga-PSMA-11
Incesu 2025 [60]	Germany	Retro, SC	1	873	873	Int/High	68Ga-PSMA-11
Cerci 2026 [61]	International	Prosp, MC	Multi	775	593	Unfav Int/High	68Ga-PSMA-11
Talavera Cobo 2026 [62]	Spain	Retro, SC	1	85	53	Int/High	68Ga-PSMA-11

Retro, retrospective; Prosp, prospective; SC, single-center; MC, multi-center; VH, very-high-risk; Int, intermediate; ePLND, extended pelvic lymph node dissection; PLND, pelvic lymph node dissection; 18F-1007, 18F-PSMA-1007.

Table 2
Study-Level 2×2 Contingency Data and Diagnostic Accuracy.

Study	TP	FP	FN	TN	N	Prev %	SE %	SP %	PPV %	NPV %	LR+	LR−
Budäus 2016	4	0	8	18	30	40.0	33.3	100.0	100.0	69.2	13.2	0.67
Maurer 2016	27	1	14	88	130	31.5	65.9	98.9	96.4	86.3	39.3	0.35
van Leeuwen 2016	7	1	4	18	30	36.7	63.6	94.7	87.5	81.8	8.3	0.41
Obek 2017	8	5	7	31	51	29.4	53.3	86.1	61.5	81.6	3.6	0.55
Zhang 2017	14	1	1	26	42	35.7	93.3	96.3	93.3	96.3	16.9	0.10
Yaxley 2019	20	9	32	147	208	25.0	38.5	94.2	69.0	82.1	6.4	0.65
Kroenke 2020	13	3	5	37	58	31.0	72.2	92.5	81.2	88.1	8.3	0.32
Esen 2021	8	1	7	80	96	15.6	53.3	98.8	88.9	92.0	29.0	0.48
Hope 2021	30	10	45	192	277	27.1	40.0	95.0	75.0	81.0	7.8	0.63
Malaspina 2021	27	1	4	47	79	39.2	87.1	97.9	96.4	92.2	28.1	0.15
Meijer 2021	72	34	117	534	757	25.0	38.1	94.0	67.9	82.0	6.3	0.66
Pientka 2021a	25	4	37	186	252	24.6	40.3	97.9	86.2	83.4	17.2	0.61
Pientka 2021b	7	6	10	94	117	14.5	41.2	94.0	53.8	90.4	6.5	0.62
Szigeti 2021	6	3	4	35	48	20.8	60.0	92.1	66.7	89.7	6.6	0.45
van Kalmthout 2021	17	5	24	51	97	42.3	41.5	91.1	77.3	68.0	4.3	0.65
Erdem 2022	3	2	2	42	49	10.2	60.0	95.5	60.0	95.5	10.5	0.44

TP, true positive; FP, false positive; FN, false negative; TN, true negative; Prev, nodal prevalence; SE, sensitivity; SP, specificity; PPV/NPV, positive/negative predictive value; LR+/LR−, positive/negative likelihood ratio. The pooled row presents bivariate random-effects estimates.

Study	TP	FP	FN	TN	N	Prev %	SE %	SP %	PPV %	NPV %	LR+	LR−
Hermsen 2022	16	7	14	62	99	30.3	53.3	89.9	69.6	81.6	5.0	0.52
Jakobsen 2022	7	3	19	75	104	25.0	26.9	96.2	70.0	79.8	6.3	0.76
Kopp 2022	1	2	4	32	39	12.8	20.0	94.1	33.3	88.9	3.5	0.81
Maestroni 2023	6	0	1	35	42	16.7	85.7	100.0	100.0	97.2	58.5	0.19
Di Giorgio 2024	14	2	2	60	78	20.5	87.5	96.8	87.5	96.8	21.5	0.15
Ozturk 2024	42	26	118	397	583	27.4	26.2	93.9	61.8	77.1	4.2	0.79
Patel 2024	6	1	1	22	30	23.3	85.7	95.7	85.7	95.7	13.0	0.20
Rosales 2024	19	11	6	53	89	28.1	76.0	82.8	63.3	89.8	4.2	0.30
Rovera 2024	12	2	1	17	32	40.6	92.3	89.5	85.7	94.4	7.1	0.12
Arena 2025	9	12	16	119	156	16.0	36.0	90.8	42.9	88.1	3.9	0.70
Esen 2025	44	24	54	399	521	18.8	44.9	94.3	64.7	88.1	7.8	0.58
Incesu 2025	100	48	120	605	873	25.2	45.5	92.6	67.6	83.4	6.1	0.59
Cerci 2026	209	23	30	331	593	40.3	87.4	93.5	90.1	91.7	13.2	0.14
Talavera Cobo 2026	11	1	1	40	53	22.6	91.7	97.6	91.7	97.6	24.8	0.12
Pooled (RE)	—	—	—	—	**5,927**	**25.5**	**56.9**	**93.2**	**83.4**	**89.0**	**7.44**	**0.50**

TP, true positive; FP, false positive; FN, false negative; TN, true negative; Prev, nodal prevalence; SE, sensitivity; SP, specificity; PPV/NPV, positive/negative predictive value; LR+/LR−, positive/negative likelihood ratio. The pooled row presents bivariate random-effects estimates.

3.3 Methodological Quality

Every study carried at least one QUADAS-2 domain rated high or unclear (Fig. 2). Patient selection was high risk in 15 studies (50%), chiefly owing to retrospective, non-consecutive enrollment; index-test interpretation was unclear in 21 (70%) because blinding to histopathology was unreported; the reference standard was low

risk in 28 (93%); and flow and timing was adequate in 28 (93%). Thirteen studies (43%) met all four domains at low risk and were designated high quality, while 17 carried one or more caveats. Restricting the synthesis to the 13 high-quality studies yielded a sensitivity of 45.7% and specificity of 93.7% (Section 3.7), indicating that lower-quality studies tended to report higher apparent sensitivity. Domain-level and study-level judgments are tabulated in Supplementary Table S2.

3.4 Pooled Diagnostic Accuracy

Across 30 studies, the bivariate random-effects model yielded a pooled sensitivity of 56.9% (95% CI 48.3–65.0%; I^2 = 86.9%, τ^2 = 0.675, Q = 220.99; $p < 0.0001$) and a pooled specificity of 93.2% (95% CI 92.1–94.2%; I^2 = 26.0%, τ^2 = 0.046, Q = 39.19; p = 0.098) (Table 3). The diagnostic odds ratio was 19.03 (95% CI 12.82–28.23), the positive predictive value 83.4% (95% CI 77.2–88.2%), and the negative predictive value 89.0% (95% CI 86.8–90.8%). Pooled likelihood ratios were 7.44 (positive; 95% CI 6.07–9.12) and 0.50 (negative; 95% CI 0.43–0.57). The Moses–Shapiro–Littenberg SROC curve had an area under the curve of 0.90 and a Q* index of 0.895, indicating strong overall discrimination.

Table 3
Pooled Diagnostic Accuracy from the Bivariate Random-Effects Model.

Parameter	Pooled Estimate	95% CI	I^2 %	τ^2	Q	p (het.)
Sensitivity	**56.9%**	**48.3–65.0%**	**86.9**	**0.675**	**220.99**	**< 0.0001**
Specificity	**93.2%**	**92.1–94.2%**	**26.0**	**0.046**	**39.19**	**0.098**
Positive predictive value	83.4%	77.2–88.2%	—	—	—	—
Negative predictive value	89.0%	86.8–90.8%	—	—	—	—
Diagnostic odds ratio	19.03	12.82–28.23	74.1	0.699	112.17	< 0.0001
Positive likelihood ratio	7.44	6.07–9.12	—	—	—	—
Negative likelihood ratio	0.50	0.43–0.57	—	—	—	—
Disease prevalence	25.5%	22.9–28.3%	—	—	—	—
Q* index (SROC)	0.895	—	—	—	—	—
SROC AUC	0.90	—	—	—	—	—
Sensitivity prediction interval	20.2–87.2%	new study	—	—	—	—
Specificity prediction interval	89.8–95.6%	new study	—	—	—	—

I^2, proportion of variation due to heterogeneity; τ^2, between-study variance; Q, Cochran's Q; SROC, summary receiver operating characteristic; AUC, area under the curve; CI, confidence interval.

The most informative contrast concerned the prediction intervals: the 95% prediction interval for sensitivity in a future study was wide (20.2–87.2%), whereas that for specificity was narrow (89.8–95.6%). This dissociation indicates that specificity is a reproducible, near-intrinsic property of the test, while sensitivity is strongly contingent on the population and methodology of the next study. Study-level estimates and pooled summaries are displayed in the sensitivity and specificity forest plots (Fig. 3 and Fig. 4), and the joint summary operating point with its 95% confidence and prediction regions on the SROC curve (Fig. 5).

3.5 Subgroup Analyses

Tracer type. 68Ga-PSMA-11 produced the highest pooled sensitivity (65.1%, 95% CI 51.5–76.6%), followed by other 68Ga agents (56.1%), 18F-PSMA-1007 (48.3%), and 18F-DCFPyL (34.0%); the single 18F-rhPSMA-7 study reported 72.2%, and mixed-tracer cohorts 33.0%. Specificity was uniformly high (90.6–94.9%) and did not differ materially across tracers. Sensitivity and specificity by tracer are compared in Fig. 6 and tabulated in Table 4.

Table 4
Subgroup Meta-Analyses by Tracer, Design, Center, Sample Size, and Verification.

Subgroup	Category	n	Sensitivity (95% CI)	Specificity (95% CI)	DOR (95% CI)	I^2 SE/SP/DOR (%)
A. By Tracer Type						
Tracer	68Ga-PSMA-11	15	65.1% (51.5–76.6)	92.9% (91.1–94.3)	25.2 (13.5–46.9)	89 / 29 / 78
Tracer	68Ga-PSMA (other)	6	56.1% (40.6–70.5)	94.0% (89.2–96.7)	24.2 (8.3–71.1)	66 / 38 / 62
Tracer	18F-PSMA-1007	5	48.3% (31.4–65.6)	93.0% (90.3–94.9)	11.4 (4.5–28.8)	83 / 25 / 51
Tracer	18F-DCFPyL	3	34.0% (23.8–46.1)	94.9% (91.5–97.0)	10.8 (4.6–25.3)	61 / 50 / 51
Tracer	18F-rhPSMA-7	1	72.2% (50.6–87.0)	92.5% (80.8–97.3)	29.4 (6.9–126.4)	— / — / —
Tracer	Mixed tracers	3	33.0% (24.7–42.6)	93.4% (91.7–94.7)	7.0 (4.6–10.7)	64 / 4 / 33
B. By Study Design and Center Type						
Design	Prospective	10	63.5% (44.9–78.8)	93.4% (91.4–95.0)	24.5 (11.1–54.1)	91 / 17 / 79
Design	Retrospective	20	51.9% (43.8–59.9)	93.2% (91.7–94.4)	14.8 (10.0–21.9)	76 / 33 / 59
Center	Single-center	24	57.3% (48.7–65.5)	92.9% (91.5–94.1)	18.3 (11.5–29.0)	78 / 19 / 72
Center	Multi-center	6	52.1% (29.8–73.6)	93.9% (91.7–95.6)	17.4 (7.4–41.0)	96 / 47 / 76
C. By Sample Size and Verification Status						
Sample size	N < 50	7	63.6% (42.5–80.5)	94.4% (90.0–96.9)	33.2 (12.0–91.7)	55 / 0 / 20

All subgroup analyses used the bivariate random-effects model. DOR, diagnostic odds ratio; SE, sensitivity; SP, specificity. The verification contrast (shaded) is the dominant methodological modifier of sensitivity. Geographic, period, risk-group, reference-standard, data-source, and image-analysis subgroups appear in Supplementary Table S3.

Subgroup	Category	n	Sensitivity (95% CI)	Specificity (95% CI)	DOR (95% CI)	I² SE/SP/DOR (%)
Sample size	N 50–99	9	70.8% (58.3–80.8)	92.4% (88.5–95.1)	35.0 (14.6–83.9)	55 / 27 / 60
Sample size	N 100–199	6	48.3% (33.9–62.9)	92.0% (86.9–95.3)	11.1 (5.6–22.2)	72 / 62 / 49
Sample size	N ≥ 200	8	46.3% (31.9–61.3)	93.8% (92.8–94.6)	14.6 (7.8–27.5)	95 / 0 / 89
Verification	All verified	25	49.4% (43.0–55.9)	93.6% (92.6–94.4)	14.5 (10.4–20.3)	72 / 6 / 56
Verification	Partial verification	5	81.7% (70.0–89.5)	90.6% (83.9–94.7)	43.5 (14.4–131.5)	51 / 59 / 71

All subgroup analyses used the bivariate random-effects model. DOR, diagnostic odds ratio; SE, sensitivity; SP, specificity. The verification contrast (shaded) is the dominant methodological modifier of sensitivity. Geographic, period, risk-group, reference-standard, data-source, and image-analysis subgroups appear in Supplementary Table S3.

Design and verification. Prospective studies showed higher sensitivity (63.5%) than retrospective studies (51.9%), with comparable specificity. The most consequential contrast was verification status: studies with partial verification reported a sensitivity of 81.7% (95% CI 70.0–89.5%) versus 49.4% (95% CI 43.0–55.9%) in completely verified cohorts—a 32.3-percentage-point difference.

Sample size and other strata. Smaller studies tended toward higher sensitivity (N < 50: 63.6%; N 50–99: 70.8%) than larger ones (N ≥ 200: 46.3%), consistent with small-study effects. Among the extended strata reported in Supplementary Table S3, sensitivity was highest in the two Asian cohorts (86.9%) and lowest in the North American cohorts (40.4%), and rose across publication periods from 47.8% in 2020–2022 to 67.9% in 2023–2026; studies that incorporated SUVmax thresholds reported higher sensitivity (65.9%) than those relying on visual interpretation alone (49.0%), whereas specificity was stable (92.7–95.4%) throughout. Subgroup results by geographic region, publication period, risk group, reference standard, data source, and image-analysis method are tabulated in full in Supplementary Table S3.

3.6 Meta-Regression

Univariable meta-regression identified three significant independent predictors of sensitivity (Table 5): median PSA (β = +0.086 per unit, p = 0.017, R^2 = 36.5%), partial versus complete verification (β = +1.331, p = 0.007, R^2 = 23.6%), and prior conventional imaging (β = +1.088, p = 0.006, R^2 = 23.9%). Median PSA was also a significant predictor of the diagnostic odds ratio (p = 0.026), and validated versus reconstructed 2×2 data significantly predicted specificity (β = +0.507, p = 0.038). Disease prevalence, quality tier, and tracer class were marginally associated with sensitivity (p = 0.060–0.095), whereas publication year, sample size, and study design were not. Collectively these covariates explain a substantial share of the high sensitivity

heterogeneity while leaving residual unexplained variation, consistent with unmeasured technical and interpretive differences across studies.

Table 5
Meta-Regression of Covariates Affecting Diagnostic Performance.

Covariate	Outcome	n	Coefficient	SE	p-value	R^2	Interpretation
Log(Sample Size)	Sensitivity	30	−0.310	0.185	0.106	0.091	Not significant
Log(Sample Size)	Specificity	30	−0.036	0.122	0.774	0.003	Not significant
Publication Year	Sensitivity	30	+ 0.058	0.065	0.382	0.027	Not significant
Disease Prevalence	Sensitivity	30	+ 3.597	2.075	0.095	0.096	Marginal
Median PSA	Sensitivity	15	+ 0.086	0.033	0.017	0.365	**Significant** ✓
Median PSA	DOR	15	+ 0.086	0.035	0.026	0.327	**Significant** ✓
Partial vs All Verified	Sensitivity	30	+ 1.331	0.458	0.007	0.236	**Significant** ✓
Partial vs All Verified	Specificity	30	−0.559	0.315	0.086	0.102	Marginal
Prior Conventional Imaging	Sensitivity	30	+ 1.088	0.379	0.006	0.239	**Significant** ✓
18F-based vs 68Ga-based	Sensitivity	30	−0.716	0.406	0.083	0.103	Marginal
Prospective vs Retrospective	Sensitivity	30	+ 0.353	0.403	0.390	0.027	Not significant
Validated vs Reconstructed 2×2	Specificity	30	+ 0.507	0.237	0.038	0.145	**Significant** ✓

Univariate weighted least-squares meta-regression. Coefficients are on the logit scale (DOR on the log scale). R^2 is the proportion of between-study heterogeneity explained by each covariate. SE, standard error; DOR, diagnostic odds ratio; PSA, prostate-specific antigen. ✓ = statistically significant ($p < 0.05$).

3.7 Sensitivity Analyses and Influence Diagnostics

Pooled sensitivity remained within a narrow band across all 14 prespecified scenarios, from 43.1% (completely verified plus directly validated 2×2 data only) to 65.1% (68Ga-PSMA-11 only), while specificity was essentially invariant (92.9–94.0%) (Table 6, Fig. 7). The most conservative estimates were obtained when restricting to high-quality studies (sensitivity 45.7%) or to completely verified, validated data (43.1%), reinforcing that the lower end of the sensitivity range is the more credible figure for clinical counseling. Leave-one-out analysis changed pooled sensitivity by no more than 3.9 percentage points; Cerci 2026 [61] exerted the greatest single-study influence (sensitivity 53.0% when omitted), and no single study materially

altered specificity (full leave-one-out diagnostics in Supplementary Table S4). These analyses confirm that the principal findings are robust to study-level and methodological choices.

Table 6
Sensitivity Analyses Across Prespecified Exclusion Scenarios.

Analysis Scenario	n	Sensitivity (95% CI)	Specificity (95% CI)	DOR (95% CI)	I^2 SE/SP (%)
Primary analysis (all studies)	**30**	**56.9% (48.3–65.0)**	**93.2% (92.1–94.2)**	**19.0 (12.8–28.2)**	**87 / 26**
Excluding reconstructed 2×2 tables	18	49.0% (40.9–57.1)	93.8% (92.6–94.8)	15.7 (10.2–24.3)	78 / 15
Excluding small studies (N < 50)	23	55.5% (45.8–64.8)	93.2% (91.9–94.3)	17.5 (11.5–26.6)	82 / 24
Excluding partial-verification studies	25	49.4% (43.0–55.9)	93.6% (92.6–94.4)	14.5 (10.4–20.3)	72 / 6
High quality only (QUADAS-2 low risk)	13	45.7% (37.7–54.0)	93.7% (92.5–94.8)	13.1 (8.6–19.8)	78 / 16
Prospective studies only	10	63.5% (44.9–78.8)	93.4% (91.4–95.0)	24.5 (11.1–54.1)	91 / 17
Retrospective studies only	20	51.9% (43.8–59.9)	93.2% (91.7–94.4)	14.8 (10.0–21.9)	76 / 33
68Ga-PSMA-11 studies only	15	65.1% (51.5–76.6)	92.9% (91.1–94.3)	25.2 (13.5–46.9)	89 / 29
18F-based tracers only	6	53.7% (36.4–70.1)	94.4% (90.9–96.6)	18.9 (8.8–40.5)	79 / 40
All verified + validated 2×2 tables	14	43.1% (35.4–51.1)	94.0% (92.8–95.0)	11.9 (8.1–17.7)	65 / 8
Excluding Zhang 2017 (outlier)	29	55.0% (46.5–63.3)	93.1% (91.9–94.1)	17.4 (11.6–25.9)	86 / 25
Excluding Cerci 2026 (highest Q)	29	53.5% (46.2–60.7)	93.2% (92.1–94.1)	16.3 (11.4–23.3)	81 / 24

Each row is a distinct exclusion scenario applied to the primary 30-study dataset; the primary analysis (shaded) is the reference. All estimates derive from the bivariate random-effects model. Leave-one-out scenarios for individual influential studies are summarized in the text and Supplementary Table S4. CI, confidence interval; DOR, diagnostic odds ratio.

3.8 Cumulative Meta-Analysis

In cumulative meta-analysis ordered by publication year, specificity converged to its final value within the first ten studies and remained stable thereafter, whereas sensitivity oscillated and its prediction interval widened as larger, more recent cohorts accrued, reflecting growing recognition of heterogeneous real-world

performance (Fig. 8). The trajectory underscores that the apparent stability of early, small high-sensitivity studies was not sustained as the evidence base matured.

3.9 Publication Bias

Small-study effects were evident on all three tests: Deeks' funnel-plot asymmetry (slope 3.70, p = 0.0009; Fig. 9), Egger's regression (p < 0.0001), and Begg's rank correlation (ρ = 0.556, p = 0.0014) [31–33]. Trim-and-fill imputed two missing studies in the lower-left quadrant, consistent with selective publication of small studies reporting higher sensitivity; the adjusted pooled sensitivity shifted downward by several percentage points, in the same direction as the completely verified and high-quality sensitivity analyses. Study-level weights and contributions to Cochran's Q are detailed in Supplementary Table S5, where Cerci 2026, Ozturk 2024, and Meijer 2021 were the three largest contributors to sensitivity heterogeneity.

4. Discussion

4.1 Principal Findings

In this synthesis of 30 studies and 5,927 patients—restricted to per-patient analyses with histopathologic ePLND reference standards in treatment-naïve intermediate-to-high-risk prostate cancer—PSMA PET/CT combined high, reproducible specificity (93.2%) with moderate, highly heterogeneous sensitivity (56.9%). The diagnostic odds ratio of 19.03 and Q* of 0.895 indicate strong overall discrimination, yet the central message lies in the dissociation between a narrow specificity prediction interval (89.8–95.6%) and a very wide sensitivity prediction interval (20.2–87.2%): specificity behaves as an intrinsic, transportable property of the test, whereas sensitivity is contingent on disease burden and study methodology. Partial-verification bias, tracer choice, and pre-scan PSA emerged as the dominant modifiers of sensitivity.

4.2 Comparison with Previous Evidence

Our estimates align closely with the most rigorous prior work while sharpening the inference to a single, well-defined clinical scenario. Mazzone and colleagues, pooling lymph-node-dissection–referenced cohorts within a 99-study review, reported a sensitivity of 54% and specificity of 94%—both within the confidence intervals of our 56.9% and 93.2% [21]. The concordance is reassuring given our deliberately narrower eligibility. Mari and colleagues reported lower sensitivity (43.6%) and specificity (85.6%), differences attributable to their inclusion of composite reference standards and mixed analysis levels that admit non-histopathologic verification [20]. Bonaddio and colleagues, focusing on intermediate-risk disease, likewise found PSMA PET/CT insufficiently sensitive to exclude nodal disease in that subgroup, mirroring our intermediate-plus-high-risk estimate [15]. The convergence of independent meta-analyses on a specificity near 93–94% and a sensitivity near 50–57% now constitutes a stable evidence base for guideline deliberation, consistent with the EAU and NCCN recommendations endorsing PSMA PET/CT for primary staging while preserving the role of histologic confirmation [3, 12]; that those guidelines rest substantially on the proPSMA [9] and OSPREY [10] trials underscores the value of anchoring the present estimate to histopathology alone.

4.3 Tracer-Specific Performance

The superior sensitivity of 68Ga-PSMA-11 (65.1%) over 18F-DCFPyL (34.0%) and 18F-PSMA-1007 (48.3%) is consistent with its earlier validation, mature reader experience, and the foundational proPSMA and OSPREY datasets [9, 10]. The 18F-DCFPyL estimate matches the histopathology-anchored OSPREY Cohort A sensitivity of 40.3% and is therefore unlikely to reflect chance. The lower observed sensitivity of 18F agents should nonetheless be interpreted cautiously: 18F tracers offer a longer half-life enabling centralized production and broader geographic access, higher intrinsic positron resolution on digital systems, and—for 18F-PSMA-1007—minimal urinary excretion that may aid assessment of the obturator fossa. Direct head-to-head comparative data remain limited and are not consistently in favor of either tracer class, with a dedicated systematic review finding broadly comparable detection across 18F and 68Ga agents [63]; the single-study 18F-rhPSMA-7 signal (72.2%) requires multicenter confirmation before clinical weight is assigned.

4.4 Partial-Verification Bias

The 32.3-percentage-point sensitivity gap between partially and completely verified studies is the most clinically important methodological finding. This is a textbook manifestation of partial-verification (work-up) bias: when only PET-positive or otherwise selected patients proceed to ePLND, patients with false-negative scans—who would be counted as false negatives in a fully verified cohort—are preferentially absent from the surgical denominator, mechanically inflating the true-positive fraction and overestimating sensitivity, exactly as predicted by the classic analysis of Begg and Greenes [64]. The meta-regression coefficient of + 1.331 corresponds to an odds ratio of roughly 3.8 for higher apparent sensitivity in partially verified designs, and partial verification alone explained nearly a quarter of the between-study sensitivity variance. Reported sensitivities from observational series with selective lymphadenectomy should therefore be regarded as optimistic, and the completely verified estimate (49.4%)—together with the validated-data estimate of 43.1%—represents the more appropriate figure for patient counseling and policy.

4.5 Size Dependence and the Micrometastasis Problem

The moderate sensitivity is mechanistically explained by metastatic deposit size. Histology-correlated cohorts show that most positive nodes 5 mm or smaller are missed, that the median size of false-negative nodes is on the order of 2–3 mm, and that sensitivity roughly doubles once deposits exceed 5 mm [15, 16]. With PET spatial resolution of 4–6 mm and partial-volume effects, micrometastatic disease lies below the detection threshold of current scanners; digital silicon-photomultiplier systems improve detectability by perhaps 15–25% but cannot abolish this physical limit, and a subset of nodal deposits show genuinely low PSMA expression that no amount of spatial resolution will recover. Clinically, a negative PSMA PET/CT does not reliably exclude small-volume nodal disease and should not, in isolation, justify omission of ePLND in patients whose nomogram-based nodal probability exceeds accepted thresholds [5].

4.6 Clinical Implications

The pooled positive likelihood ratio of 7.44 raises the post-test probability of N1 disease from the baseline prevalence of 25.5% to approximately 72%, so a positive scan substantially justifies treatment intensification—extension of radiation fields, addition of systemic therapy, or tailored lymphadenectomy. The negative likelihood ratio of 0.50 lowers post-test probability only to about 15%, meaning roughly one in seven “negative” patients still harbors nodal metastasis. PSMA PET/CT is therefore best deployed as a rule-in test

that complements, rather than replaces, validated nomograms and ePLND. That said, in carefully selected lower-burden populations the high negative predictive value may have practical value: a single-center analysis found no significant difference in biochemical-recurrence-free survival between intermediate-risk patients with a negative PSMA PET who did or did not undergo PLND, suggesting that dissection might be safely omitted in a defined subset—although the small size of nodal deposits missed in that cohort means residual occult disease cannot be excluded [65]. The test's high specificity also supports redirecting care away from non-curative surgery when occult metastatic disease is detected, as demonstrated in the prospective multicenter IAEA cohort [61].

4.7 Strengths and Limitations

Strengths include stringent, scenario-specific eligibility criteria; independent dual verification of every 2×2 table against source publications; a bivariate hierarchical analytic framework with prediction intervals; comprehensive subgroup, meta-regression, sensitivity, leave-one-out, and cumulative analyses; and full adherence to PRISMA-DTA [22, 23]. Limitations include substantial residual sensitivity heterogeneity not fully explained by measured covariates—likely reflecting unrecorded inter-reader variability, SUVmax thresholds, scanner generation, and patient-preparation differences; a predominance of single-center, retrospective, European cohorts that may limit generalizability; reliance on aggregate rather than individual-patient data, precluding optimal-threshold modeling; and demonstrable publication bias favoring small high-sensitivity studies, which our completely verified and validated-data sensitivity analyses were designed to mitigate.

4.8 Future Directions

Priorities include prospective multicenter trials with mandatory ePLND in all participants to eliminate verification bias; harmonized acquisition and interpretation following the EANM/SNMMI PSMA procedure standard 2.0 with standardized SUVmax reporting [8]; cost-effectiveness analyses across health systems to inform reimbursement, building on the proPSMA-derived evidence that PSMA PET/CT can be cost-dominant where scan costs are competitive [66]; validated artificial-intelligence lesion-detection tools tested specifically in primary nodal staging, where current reviews report high accuracy but variable sensitivity that still requires expert oversight [18, 67]; head-to-head tracer comparisons [63] and dedicated evaluation of combined PSMA PET/MRI for small-node detection [19]; and integration with maturing real-world analyses of PSMA staging in intermediate-risk disease [68]. The ongoing randomized evaluation of PSMA-guided versus standard lymphadenectomy will be particularly informative once per-patient accuracy data mature [17].

5. Conclusion

Pooling the most rigorously selected evidence to date, this meta-analysis shows that PSMA PET/CT is a powerful rule-in test but an unreliable rule-out test for pelvic nodal disease in treatment-naïve intermediate-to-high-risk prostate cancer. Its specificity (93.2%) is high and remarkably reproducible across populations, so a positive scan—raising the probability of nodal metastasis from roughly one in four to nearly three in four—confidently identifies men who warrant treatment intensification and can spare some patients non-curative surgery. Its sensitivity, by contrast, is only moderate (56.9%) and profoundly methodology-dependent: once the inflation introduced by partial verification is removed, the credible figure falls to the 43–49% range, and roughly one in seven men with a negative scan still harbors histologic nodal disease.

This asymmetry is not a statistical artifact but a physical consequence of a detection threshold that sub-5-mm micrometastases routinely defeat. The clinical corollary is therefore unambiguous: a negative PSMA PET/CT must not, on its own, be used to withhold extended pelvic lymph node dissection from a man whose nomogram-predicted nodal risk is high; rather, the scan should be integrated with validated nomograms, tracer- and protocol-aware interpretation, and selective lymphadenectomy. These findings support—and delimit—current guideline endorsement of PSMA PET/CT for primary staging, and they define a concrete research agenda: fully verified prospective trials, harmonized acquisition and reporting, validated artificial-intelligence assistance, and head-to-head tracer comparisons that together can extend the reliable boundaries of this transformative technology.

Declarations

Ethics approval and consent to participate.

Not applicable. This review does not involve human subjects, animal studies, or individual human data requiring ethics approval.

Consent for publication.

Not applicable.

Conflict of interest.

The authors declare no conflict of interest.

Clinical trial number

Not applicable

Funding.

This research received no specific grant from any public, commercial, or not-for-profit funding agency.

Author Contribution

B.M. and R.S.D. conceived and designed the study, developed the protocol, and drafted the manuscript. B.M., R.S.D., and E.Z. conducted the systematic literature search, screening, and full-text review. B.M., R.S.D., E.Z., and N.N. independently extracted data and verified all 2×2 contingency tables against source publications. B.M. and R.S.D. performed the statistical analyses, including bivariate random-effects modeling, subgroup analyses, meta-regression, sensitivity analyses, and publication-bias assessment. M.M.A., H.M., and S.F. contributed to data interpretation and critically revised the manuscript. A.N., K.A., and H.N. provided methodological oversight, contributed to the GRADE-informed certainty assessment, and critically revised the manuscript for important intellectual content. R.S.D. is the corresponding author and takes responsibility

for the integrity of the work from inception to publication. All authors approved the final manuscript and agree to be accountable for all aspects of the work, ensuring that questions related to accuracy or integrity are appropriately investigated and resolved.

Data Availability

All data analyzed are contained within the article and its Supplementary Material; extraction files are available from the corresponding author on reasonable request.

References

1. Siegel RL, Kratzer TB, Giaquinto AN et al (2025) Cancer statistics, 2025. CA Cancer J Clin 75(1):10–45. 10.3322/caac.21871
2. Kratzer TB, Mazzitelli N, Star J, Dahut WL, Jemal A, Siegel RL (2025) Prostate cancer statistics, 2025. CA Cancer J Clin 75(6):485–497. 10.3322/caac.70028
3. Cornford P, van den Bergh RCN, Briers E et al (2024) EAU-EANM-ESTRO-ESUR-ISUP-SIOG Guidelines on Prostate Cancer—2024 Update. Part I. Eur Urol 86(2):148–163. 10.1016/j.eururo.2024.03.027
4. Mottet N, van den Bergh RCN, Briers E et al (2021) EAU-EANM-ESTRO-ESUR-ISUP-SIOG Guidelines on Prostate Cancer. Eur Urol 79(2):243–262. 10.1016/j.eururo.2020.09.042
5. Morizane S, Takenaka A (2024) Current status and therapeutic value of extended pelvic lymph node dissection during radical prostatectomy for prostate cancer. Prostate Int 12(3):117–127. 10.1016/j.prnil.2024.06.001
6. Hövels AM, Heesakkers RA, Adang EM et al (2008) The diagnostic accuracy of CT and MRI in the staging of pelvic lymph nodes in prostate cancer: a meta-analysis. Clin Radiol 63(4):387–395. 10.1016/j.crad.2007.05.022
7. Ghosh A, Heston WD (2004) Tumor target prostate specific membrane antigen (PSMA) and its regulation in prostate cancer. J Cell Biochem 91(3):528–539. 10.1002/jcb.10661
8. Fendler WP, Eiber M, Beheshti M et al (2023) PSMA PET/CT: joint EANM procedure guideline/SNMMI procedure standard for prostate cancer imaging 2.0. Eur J Nucl Med Mol Imaging 50(5):1466–1486. 10.1007/s00259-022-06089-w
9. Hofman MS, Lawrentschuk N, Francis RJ et al (2020) Prostate-specific membrane antigen PET-CT in patients with high-risk prostate cancer before curative-intent surgery or radiotherapy (proPSMA): a prospective, randomised, multicentre study. Lancet 395(10231):1208–1216. 10.1016/S0140-6736(20)30314-7
10. Pienta KJ, Gorin MA, Rowe SP et al (2021) A phase 2/3 prospective multicenter study of the diagnostic accuracy of prostate-specific membrane antigen PET/CT with 18F-DCFPyL in prostate cancer patients (OSPREY). J Urol 206(1):52–61. 10.1097/JU.0000000000001698
11. Sartor O, de Bono J, Chi KN et al (2021) Lutetium-177-PSMA-617 for metastatic castration-resistant prostate cancer (VISION). N Engl J Med 385(12):1091–1103. 10.1056/NEJMoa2107322

12. National Comprehensive Cancer Network (2024) NCCN Clinical Practice Guidelines in Oncology: Prostate Cancer. Version 1.2025. NCCN, Plymouth Meeting, PA
13. Carll J, Shi W, Perera M, Lawrentschuk N, Chengodu T, Woon D (2025) Guideline of guidelines: PSMA PET in staging newly diagnosed intermediate-risk prostate cancer. BJU Int 136(5):800–804. 10.1111/bju.16872
14. Lowrance W, Dreicer R, Jarrard DF et al (2023) Updates to advanced prostate cancer: AUA/SUO guideline (2023). J Urol 209(6):1082–1090. 10.1097/JU.0000000000003452
15. Bonaddio J, Carll J, Eapen R, Sathianathen N, Lawrentschuk N (2026) The role of PSMA PET/CT in staging patients with intermediate-risk prostate cancer: a systematic review and diagnostic meta-analysis. BJU Int 137(1):49–57. 10.1111/bju.70015
16. Duan H, Moradi F, Davidzon GA et al (2024) PSMA PET for nodal staging: impact of lesion size and SUVmax thresholds. J Nucl Med 65(2):201–208
17. Jia X, Wang L, Zhang Y et al (2026) Prospective randomized study of 18F-PSMA PET/CT-guided pelvic lymph node dissection in prostate cancer. Prostate Cancer Prostatic Dis. 10.1038/s41391-025-00938-3
18. Liu J, Cundy TP, Woon DTS, Lawrentschuk N (2024) A systematic review on artificial intelligence evaluating metastatic prostatic cancer and lymph nodes on PSMA PET scans. Cancers (Basel) 16(3):486. 10.3390/cancers16030486
19. Moradi F, Duan H, Song H et al (2022) 68Ga-PSMA-11 PET/MRI in patients with newly diagnosed intermediate- or high-risk prostate adenocarcinoma. J Nucl Med 63(12):1822–1828. 10.2967/jnumed.121.263692
20. Mari A, Cadenar AC, Giudici S et al (2025) A systematic review and meta-analysis to evaluate the diagnostic accuracy of PSMA PET/CT in the initial staging of prostate cancer. Prostate Cancer Prostatic Dis 28(1):56–69. 10.1038/s41391-024-00850-y
21. Mazzone E, Cannoletta D, Quarta L et al (2025) A comprehensive systematic review and meta-analysis of the role of prostate-specific membrane antigen positron emission tomography for prostate cancer diagnosis and primary staging before definitive treatment. Eur Urol 87(6):654–671. 10.1016/j.eururo.2025.03.002
22. Page MJ, McKenzie JE, Bossuyt PM et al (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372:n71. 10.1136/bmj.n71
23. McInnes MDF, Moher D, Thombs BD et al (2018) Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: the PRISMA-DTA statement. JAMA 319(4):388–396. 10.1001/jama.2017.19163
24. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y (eds) (2023) Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Version 2.0. Cochrane, London
25. Whiting PF, Rutjes AW, Westwood ME et al (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 155(8):529–536. 10.7326/0003-4819-155-8-201110180-00009
26. Reitsma JB, Glas AS, Rutjes AW et al (2005) Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 58(10):982–990.

10.1016/j.jclinepi.2005.02.022

27. Rutter CM, Gatsonis CA (2001) A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med 20(19):2865–2884. 10.1002/sim.942
28. DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. Control Clin Trials 7(3):177–188. 10.1016/0197-2456(86)90046-2
29. Moses LE, Shapiro D, Littenberg B (1993) Combining independent studies of a diagnostic test into a summary ROC curve. Stat Med 12(14):1293–1316. 10.1002/sim.4780121403
30. Higgins JPT, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. BMJ 327(7414):557–560. 10.1136/bmj.327.7414.557
31. Deeks JJ, Macaskill P, Irwig L (2005) The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 58(9):882–893. 10.1016/j.jclinepi.2005.01.016
32. Egger M, Davey Smith G, Schneider M, Minder C (1997) Bias in meta-analysis detected by a simple, graphical test. BMJ 315(7109):629–634. 10.1136/bmj.315.7109.629
33. Begg CB, Mazumdar M (1994) Operating characteristics of a rank correlation test for publication bias. Biometrics 50(4):1088–1101
34. Budäus L, Leyh-Bannurah SR, Salomon G et al (2016) Initial experience of 68Ga-PSMA PET/CT imaging in high-risk prostate cancer patients prior to radical prostatectomy. Eur Urol 69(3):393–396. 10.1016/j.eururo.2015.06.027
35. Maurer T, Gschwend JE, Rauscher I et al (2016) Diagnostic efficacy of 68gallium-PSMA positron emission tomography compared to conventional imaging for lymph node staging of 130 consecutive patients with intermediate to high risk prostate cancer. J Urol 195(5):1436–1443. 10.1016/j.juro.2015.12.025
36. van Leeuwen PJ, Emmett L, Ho B et al (2017) Prospective evaluation of 68gallium-prostate-specific membrane antigen positron emission tomography in advanced prostate cancer. BJU Int 119(2):209–215. 10.1111/bju.13619
37. Obek C, Doganca T, Demirci E et al (2017) The accuracy of 68Ga-PSMA PET/CT in primary lymph node staging in high-risk prostate cancer. Eur J Nucl Med Mol Imaging 44(11):1806–1812. 10.1007/s00259-017-3752-y
38. Zhang Q, Zang S, Zhang C et al (2017) Comparison of 68Ga-PSMA-11 PET-CT with mpMRI for preoperative lymph node staging in patients with intermediate to high-risk prostate cancer. J Transl Med 15(1):230. 10.1186/s12967-017-1333-2
39. Yaxley JW, Raveenthiran S, Nouhaud FX et al (2019) Outcomes of primary lymph node staging of intermediate and high risk prostate cancer with 68Ga-PSMA PET/CT compared to histological correlation of pelvic lymph node pathology. J Urol 201(4):815–820. 10.1097/JU.0000000000000053
40. Kroenke M, Rauscher I, Horn T et al (2020) Histologically confirmed diagnostic efficacy of 18F-rhPSMA-7 PET for N-staging of patients with primary high-risk prostate cancer. J Nucl Med 61(3):429–434. 10.2967/jnumed.119.234906

41. Esen T, Falay O, Tarim K et al (2021) 68Ga-PSMA-11 PET/CT for primary lymph node staging before radical prostatectomy: central review and comparison with histopathology of extended lymphadenectomy. Eur Urol Focus 7(2):288–293. 10.1016/j.euf.2020.12.003
42. Hope TA, Eiber M, Armstrong WR et al (2021) Diagnostic accuracy of 68Ga-PSMA-11 PET/CT for pelvic nodal metastasis detection prior to radical prostatectomy and pelvic lymph node dissection: a multicenter prospective phase 3 imaging trial. JAMA Oncol 7(11):1635–1642. 10.1001/jamaoncol.2021.3771
43. Malaspina S, Anttinen M, Taimen P et al (2021) Prospective comparison of 18F-PSMA-1007 PET/CT, whole-body MRI and CT in primary nodal staging of unfavourable intermediate- and high-risk prostate cancer. Eur J Nucl Med Mol Imaging 48(9):2951–2959. 10.1007/s00259-021-05296-1
44. Meijer D, van Leeuwen PJ, Roberts MJ et al (2021) Tumor-positive pelvic lymph node identification and risk stratification using 68Ga-PSMA-11 PET/CT in primary staging of prostate cancer. Eur J Nucl Med Mol Imaging 48(6):1939–1950. 10.1007/s00259-020-05117-4
45. Jansen BHE, Bodar YJL, Zwezerijnen GJC et al (2021) Pelvic lymph-node staging with 18F-DCFPyL PET/CT prior to extended pelvic lymph-node dissection in primary prostate cancer—the SALT trial. Eur J Nucl Med Mol Imaging 48(2):509–520. 10.1007/s00259-020-04944-9
46. Szigeti F, Schweighofer-Zwink G, Meissnitzer M et al (2022) Incremental impact of 68Ga-PSMA-11 PET/CT in primary N and M staging of prostate cancer prior to curative-intent surgery: a prospective clinical trial in comparison with mpMRI. Mol Imaging Biol 24(1):50–59. 10.1007/s11307-021-01650-9
47. van Kalmthout LWM, van Melick HHE, Dijkstra JR et al (2021) Prospective validation of gallium-68 prostate-specific membrane antigen positron emission tomography/computed tomography for primary staging of prostate cancer. J Urol 206(3):561–568. 10.1097/JU.0000000000001771
48. Erdem S, Simsek DH, Degirmenci E et al (2022) How accurate is 68Ga-PSMA PET/CT on primary lymph node staging before radical prostatectomy in intermediate and high risk prostate cancer? A study of patient- and lymph node-based analyses. Urol Oncol 40(1):6. 10.1016/j.urolonc.2021.07.008
49. Hermsen R, Wedick EBC, Vinken MJM et al (2022) Lymph node staging with 18F-PSMA-1007 PET/CT in newly diagnosed intermediate- to high-risk prostate cancer using histopathological evaluation of extended pelvic node dissection as reference. Eur J Nucl Med Mol Imaging 49(11):3929–3937. 10.1007/s00259-022-05827-4
50. Jakobsen JS, Trölle A, Graversen P et al (2022) Diagnostic accuracy of 18F-PSMA-1007 PET/CT for lymph node staging in primary staging of prostate cancer. Urol Oncol 40(8):373. e1-373.e8
51. Kopp D, Kopp J, Bernhardt E et al (2022) 68Ga-PSMA PET/CT-based primary staging and histological correlation after extended pelvic lymph node dissection in intermediate-risk prostate cancer. Urol Int 106(1):56–62. 10.1159/000517834
52. Maestroni UV, Campobasso D, Guarino G et al (2023) Lymph node staging with 68Ga-PSMA PET in patients with intermediate and high-risk prostate cancer suitable for radical prostatectomy managed in a prostate cancer unit. Chin Clin Oncol 12(3):22. 10.21037/cco-23-22
53. Di Giorgio A, Rapa M, Civollani S, Farolfı A, Fanti S (2024) Diagnostic accuracy of 68Ga-PSMA PET/CT in lymph node staging of intermediate and high-risk prostate cancer patients. Curr Oncol 32(1):9. 10.3390/curroncol32010009

54. Ozturk H, Cicek E, Kose E et al (2024) Diagnostic accuracy of 68Ga-PSMA-11 and 18F-DCFPyL PET/CT for lymph node staging in intermediate- and high-risk prostate cancer: a single-center experience. Ann Nucl Med 38(9):710–719. 10.1007/s12149-024-01950-4
55. Patel KS, Singh T, Raghuvanshi K, Sonar S, Chaudhari R (2024) A comparison study of 68gallium-PSMA PET/CT and multiparametric MRI for locoregional staging of prostate cancer. Urol Sci 35(1):36–41. 10.1097/US9.0000000000000002
56. Rosales JJ, Betech-Antar V, Mínguez F et al (2025) Comparison of staging using 68Ga-PSMA-11 PET/CT and histopathological results in intermediate- and high-risk prostate cancer patients treated with radical prostatectomy and pelvic lymph node dissection. Rev Esp Med Nucl Imagen Mol (Engl Ed) 44(2):500076. 10.1016/j.remnie.2024.500076
57. Rovera G, Grimaldi S, Oderda M et al (2024) Comparative performance of 68Ga-PSMA-11 PET/CT and conventional imaging in the primary staging of high-risk prostate cancer patients who are candidates for radical prostatectomy. Diagnostics (Basel) 14(17):1964. 10.3390/diagnostics14171964
58. Arena P, Dell'Oglio P, Stabile A et al (2025) An evaluation of the diagnostic accuracy of 68Ga-PSMA-11 vs. 18F-PSMA-1007 PET/CT for lymph node staging in candidates for radical prostatectomy and lymph node dissection: a single-institutional analysis. Diagnostics (Basel) 15(12):1492. 10.3390/diagnostics15121492
59. Esen B, Seymen H, Armutlu A et al (2025) Impact of pelvic lymph node dissection on early oncological outcomes in intermediate-risk prostate cancer patients with node-negative PSMA PET. Prostate 85(8):777–783. 10.1002/pros.24884
60. Incesu RB, Preisser F, Pompe RS et al (2025) Diagnostic accuracy of prostate-specific membrane antigen positron emission tomography/computed tomography for primary lymph node staging before radical prostatectomy. Eur Urol Focus 11(3):582–587. 10.1016/j.euf.2025.01.015
61. Cerci JJ, Fanti S, Lobato EE et al (2026) 68Ga-PSMA-11 in staging of unfavorable intermediate- and high-risk prostate cancer reduces indication for noncurative prostatectomy: a prospective, multicenter, IAEA study. J Nucl Med 67(1):67–70. 10.2967/jnumed.125.270537
62. Talavera Cobo V, Yáñez Ruiz CA, Tapia Tapia MD et al (2026) Diagnostic accuracy of PSMA-PET/CT vs. mpMRI in primary staging of intermediate- and high-risk prostate cancer. Med Sci (Basel) 14(1):64. 10.3390/medsci14010064
63. Bin X, Wang L, Yang J et al (2024) Comparison of 18F-based PSMA radiotracers with 68Ga-PSMA-11 in PET/CT imaging of prostate cancer: a systematic review and meta-analysis. Prostate Cancer Prostatic Dis 27(1):44–53. 10.1038/s41391-023-00755-2
64. Begg CB, Greenes RA (1983) Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics 39(1):207–215. 10.2307/2530820
65. Incesu RB, Preisser F, Nohe F, Maurer T, Graefen M, Tilki D (2025) Negative PSMA PET can be used to avoid unnecessary pelvic lymph node dissection in intermediate-risk prostate cancer. Prostate Cancer Prostatic Dis 28(4):874–878. 10.1038/s41391-024-00930-z
66. de Feria Cardet RE, Hofman MS, Segard T et al (2021) Is prostate-specific membrane antigen positron emission tomography/computed tomography imaging cost-effective in prostate cancer: an analysis informed by the proPSMA trial. Eur Urol 79(3):413–418. 10.1016/j.eururo.2020.11.043

67. Dadgar H, Hong X, Karimzadeh R et al (2025) The value of artificial intelligence in PSMA PET: a pathway to improved efficiency and results. Q J Nucl Med Mol Imaging 69(2):157–173. 10.23736/S1824-4785.25.03640-4
68. Hartwieg BF, Maurer T, Kniep C et al (2026) Use of PSMA staging in intermediate-risk prostate cancer: a real-world data analysis. Urol Oncol 44(2):71. e1-71.e7

Figures

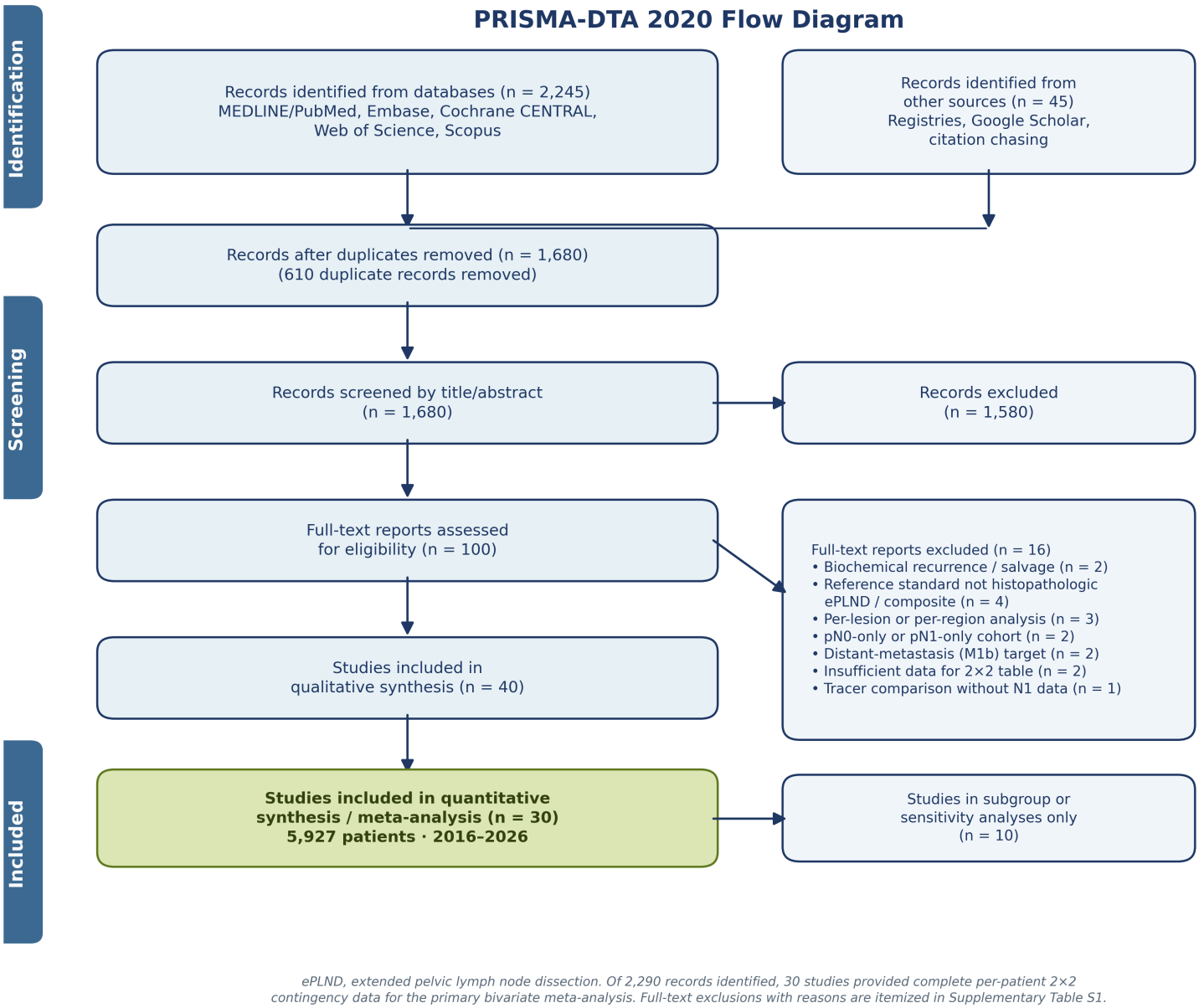


Figure 1

PRISMA-DTA 2020 flow diagram. Flow of records through identification, screening, eligibility, and inclusion. Of 2,290 identified records, 30 studies with complete per-patient 2×2 data entered the primary bivariate meta-analysis.

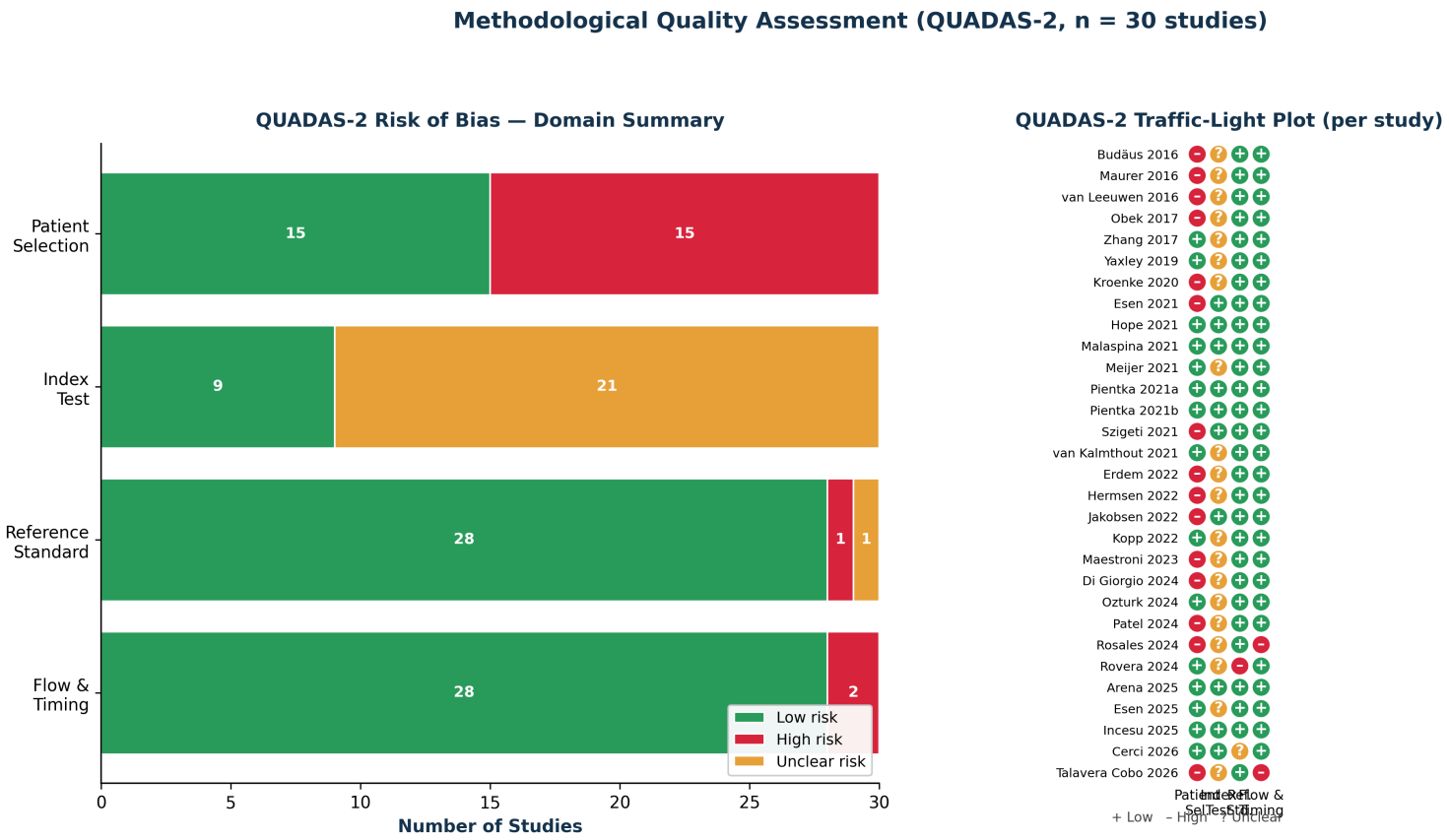


Figure 2

QUADAS-2 methodological quality. Left: stacked domain-level summary of low, high, and unclear risk of bias across the 30 studies. Right: per-study traffic-light plot. Index-test interpretation was the most frequent source of unclear risk, reflecting unreported reader blinding.

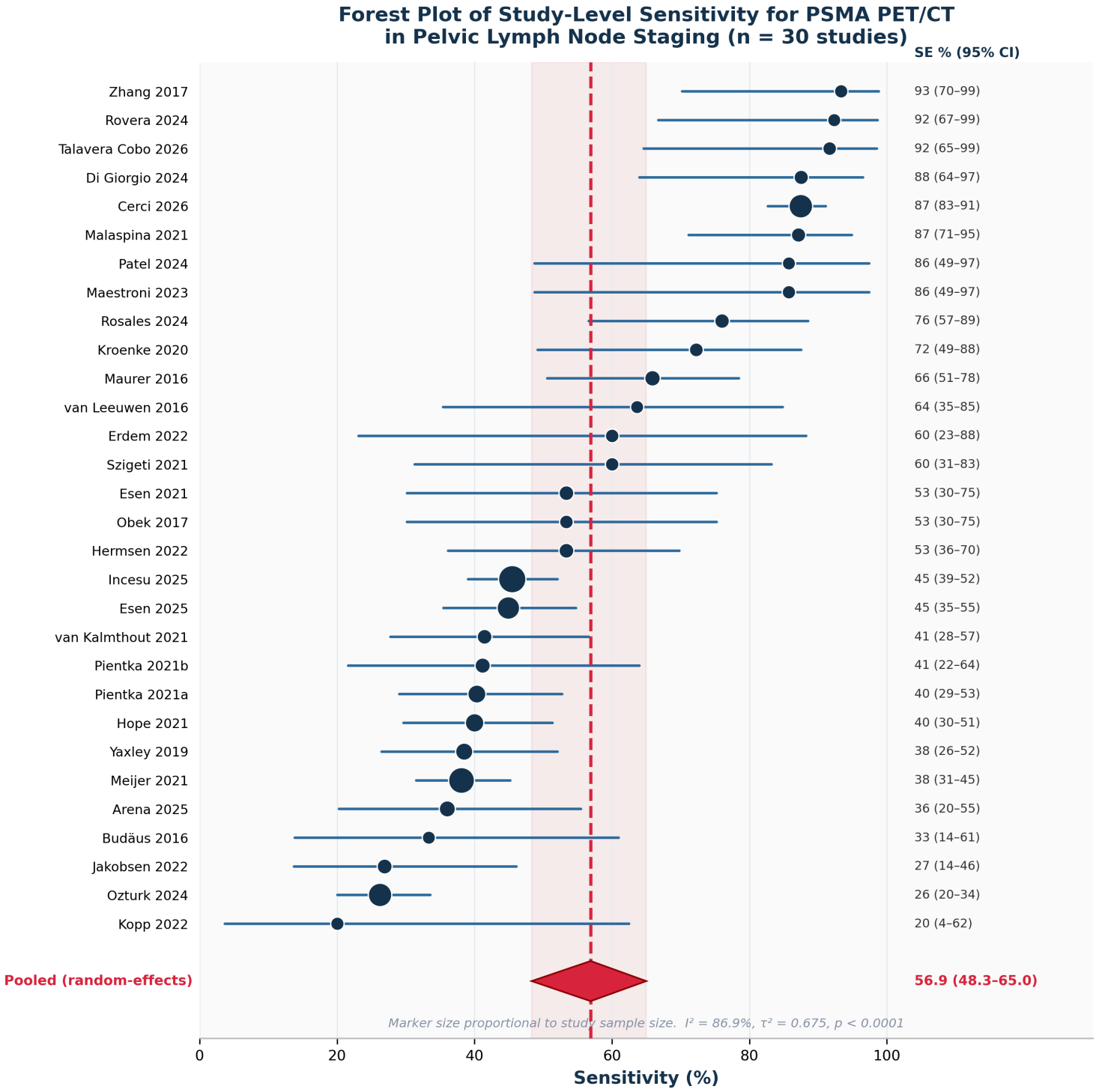


Figure 3

Forest plot of study-level sensitivity. Point estimates (markers sized in proportion to sample size) with 95% Wilson confidence intervals for each of the 30 studies, ordered by sensitivity. The diamond and shaded band denote the pooled random-effects estimate (56.9%, 95% CI 48.3–65.0%). Heterogeneity was substantial (I^2 = 86.9%).

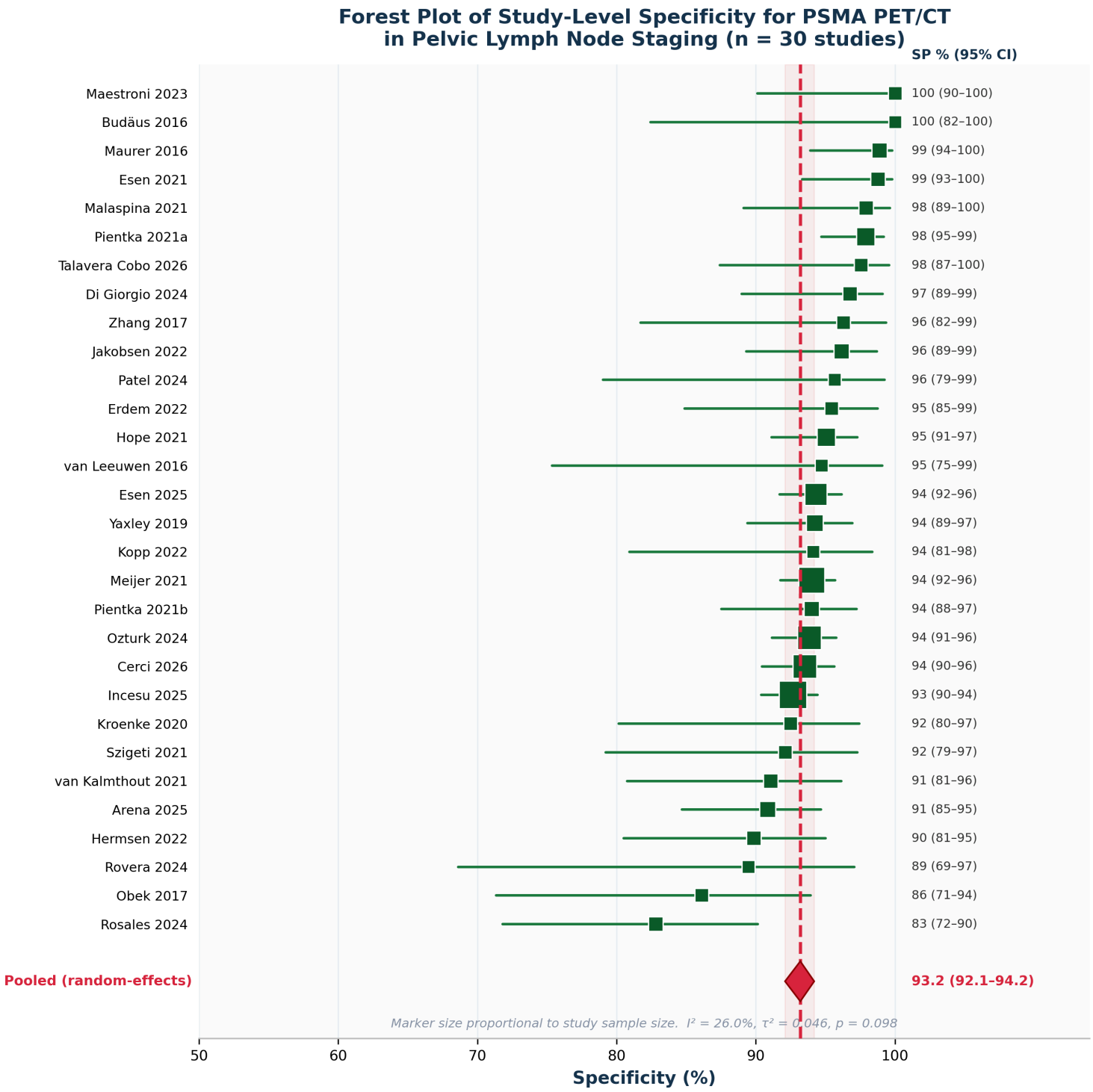


Figure 4

Forest plot of study-level specificity. Point estimates with 95% Wilson confidence intervals, ordered by specificity. The pooled specificity (93.2%, 95% CI 92.1–94.2%) is shown by the diamond and shaded band. Heterogeneity was low to moderate (I^2 = 26.0%).

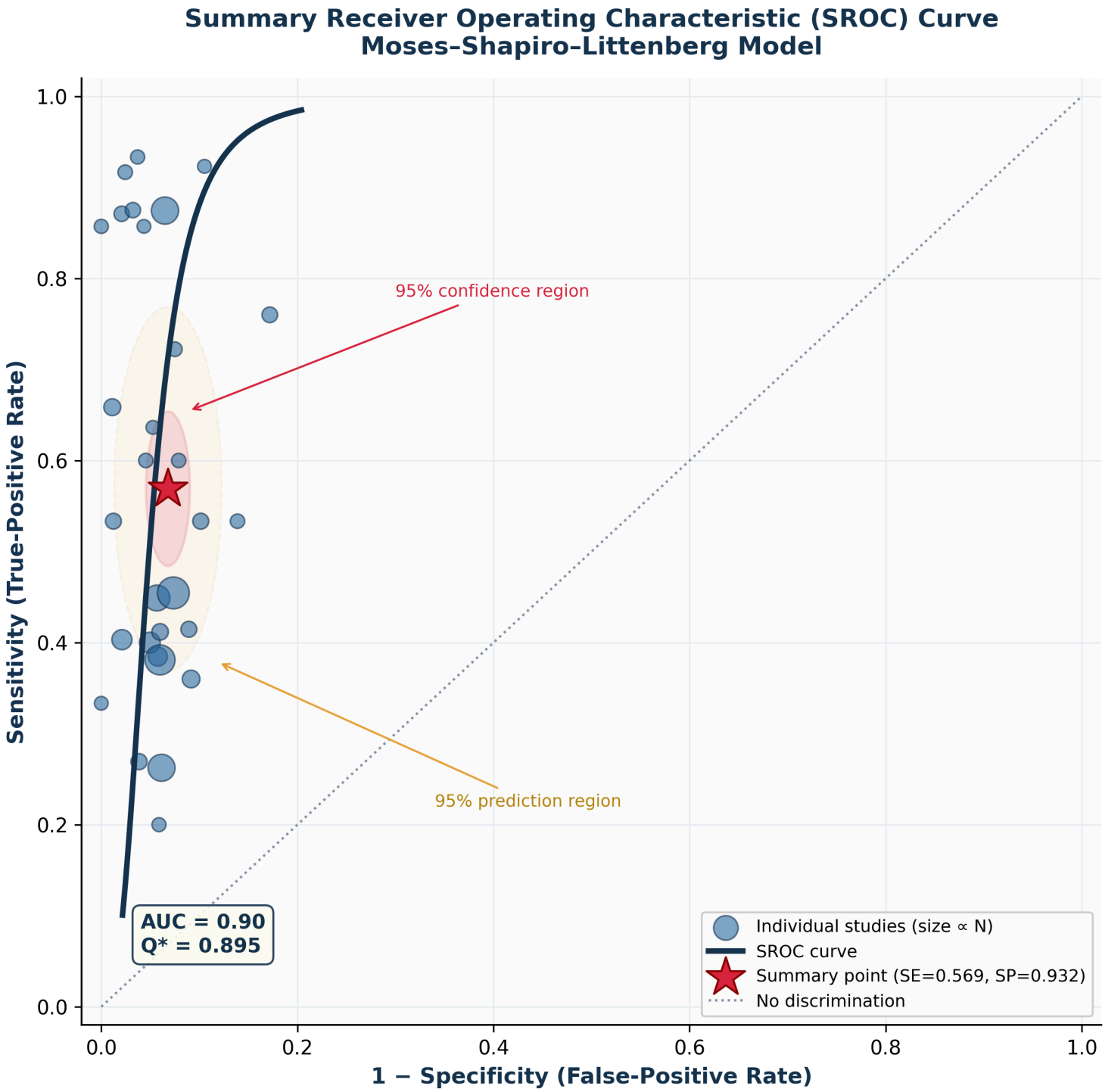


Figure 5

Summary receiver operating characteristic (SROC) curve. Moses–Shapiro–Littenberg SROC curve with individual study estimates (markers sized by sample size), the summary operating point (sensitivity 0.569, specificity 0.932), and 95% confidence and prediction regions. The area under the curve was 0.90 and the Q* index 0.895.

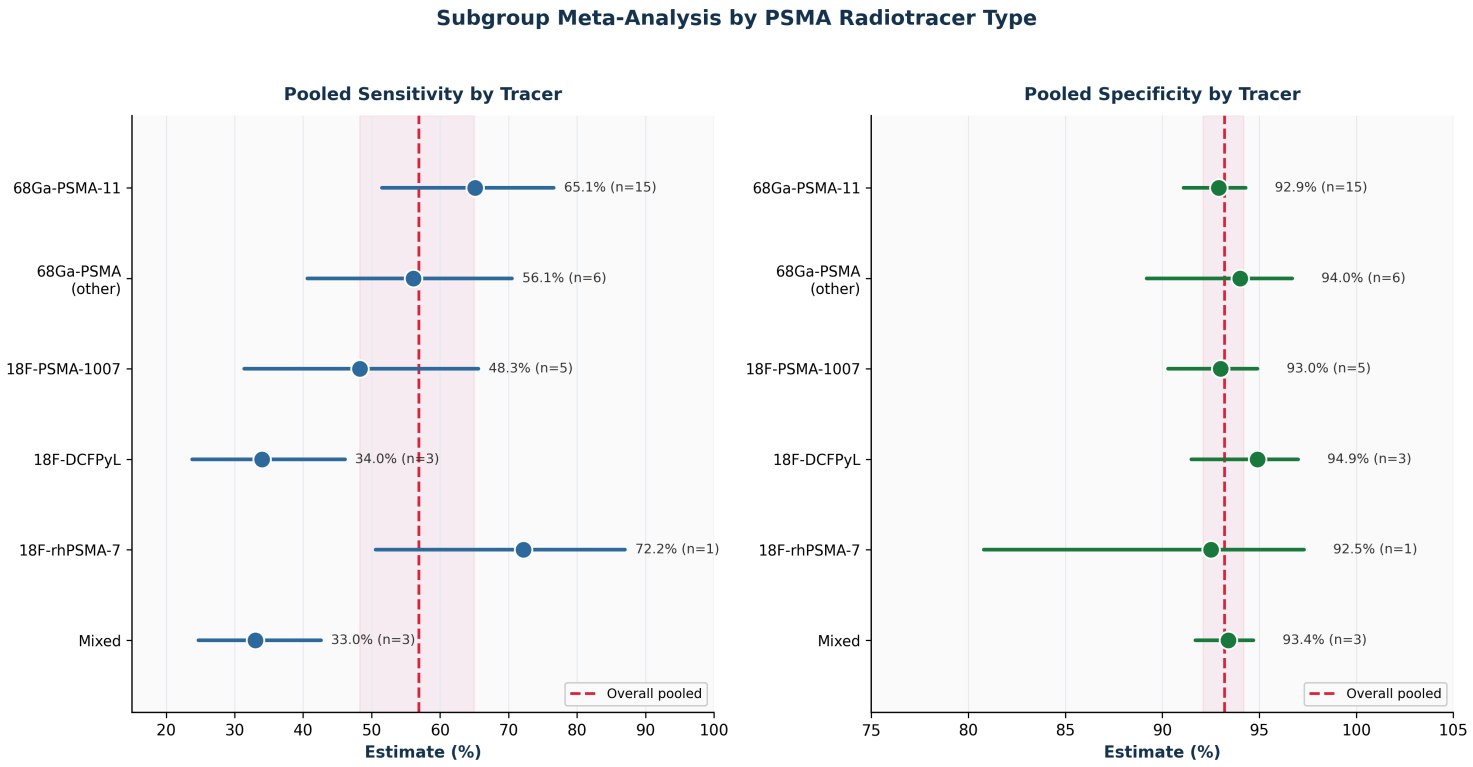


Figure 6

Subgroup meta-analysis by PSMA radiotracer. Pooled sensitivity (left) and specificity (right) with 95% confidence intervals for each tracer subgroup; the number of studies is annotated and the dashed line marks the overall pooled estimate. 68Ga-PSMA-11 showed the highest sensitivity, whereas specificity was uniformly high.

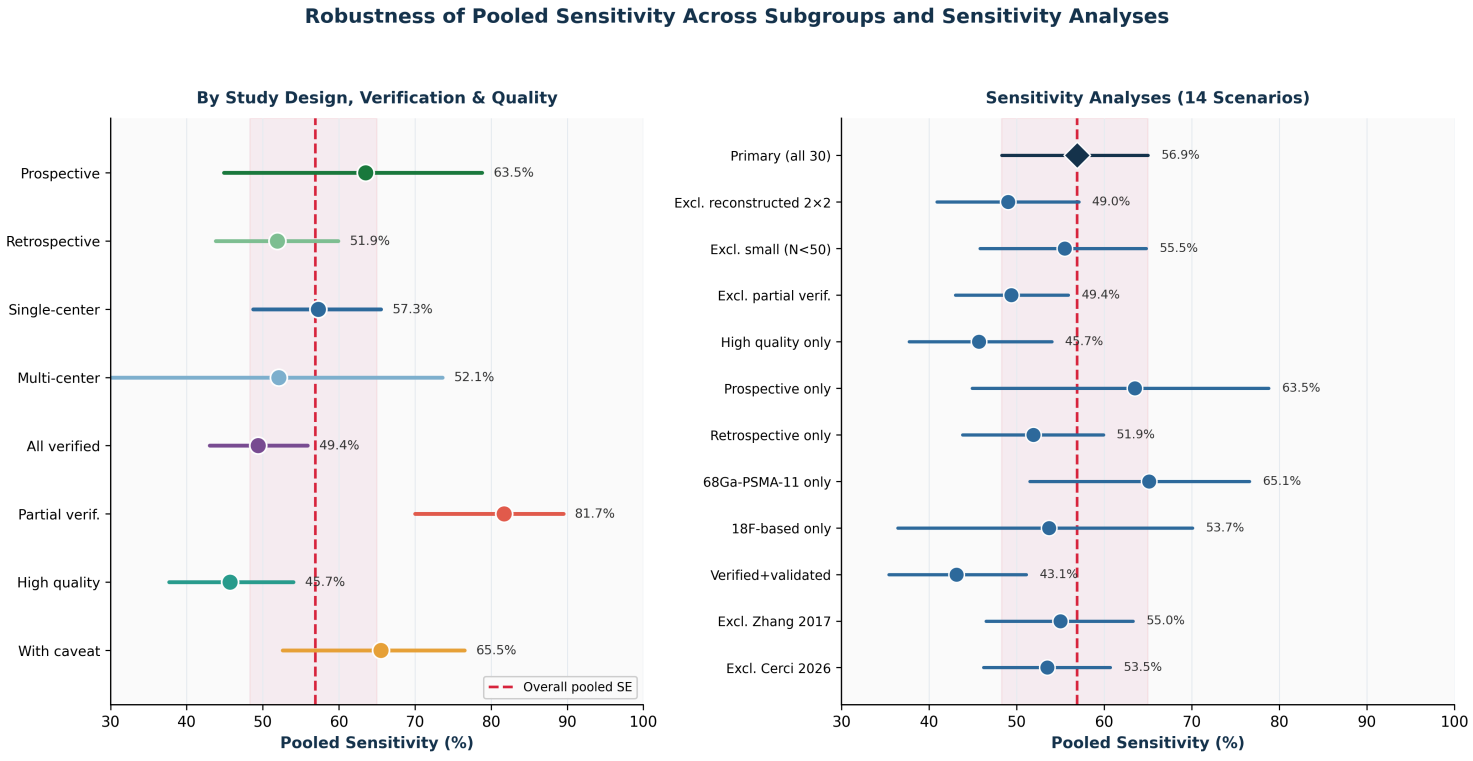


Figure 7

Subgroup and sensitivity analyses of pooled sensitivity. Left: pooled sensitivity by study design, verification status, and quality tier. Right: pooled sensitivity across the prespecified sensitivity-analysis scenarios (primary analysis shown as a diamond). Estimates remained within 43.1–65.1% throughout.

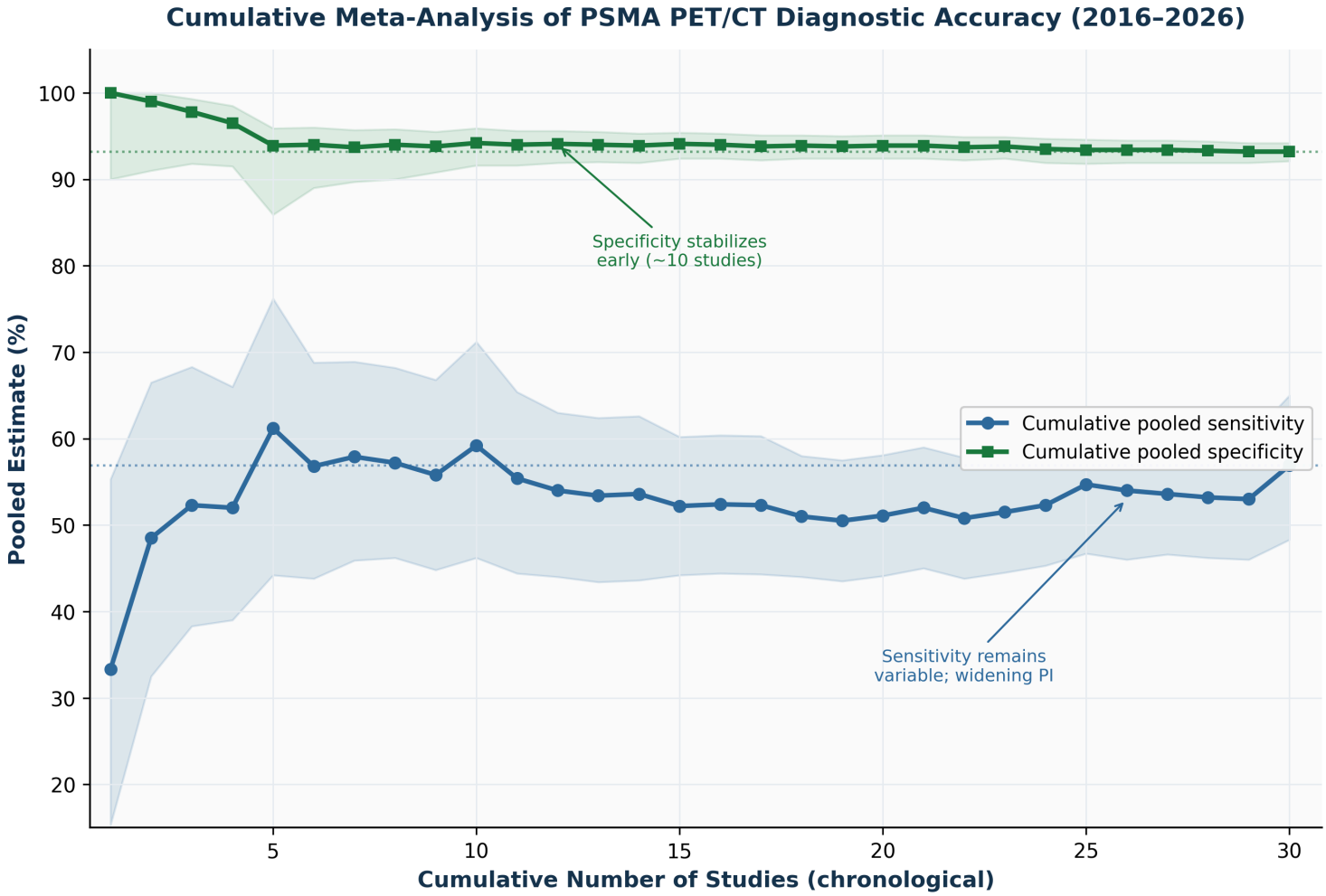


Figure 8

Cumulative meta-analysis (2016–2026). Pooled sensitivity and specificity recomputed as studies accrue chronologically, with 95% confidence bands. Specificity stabilized within ten studies; sensitivity remained variable with a widening prediction interval.

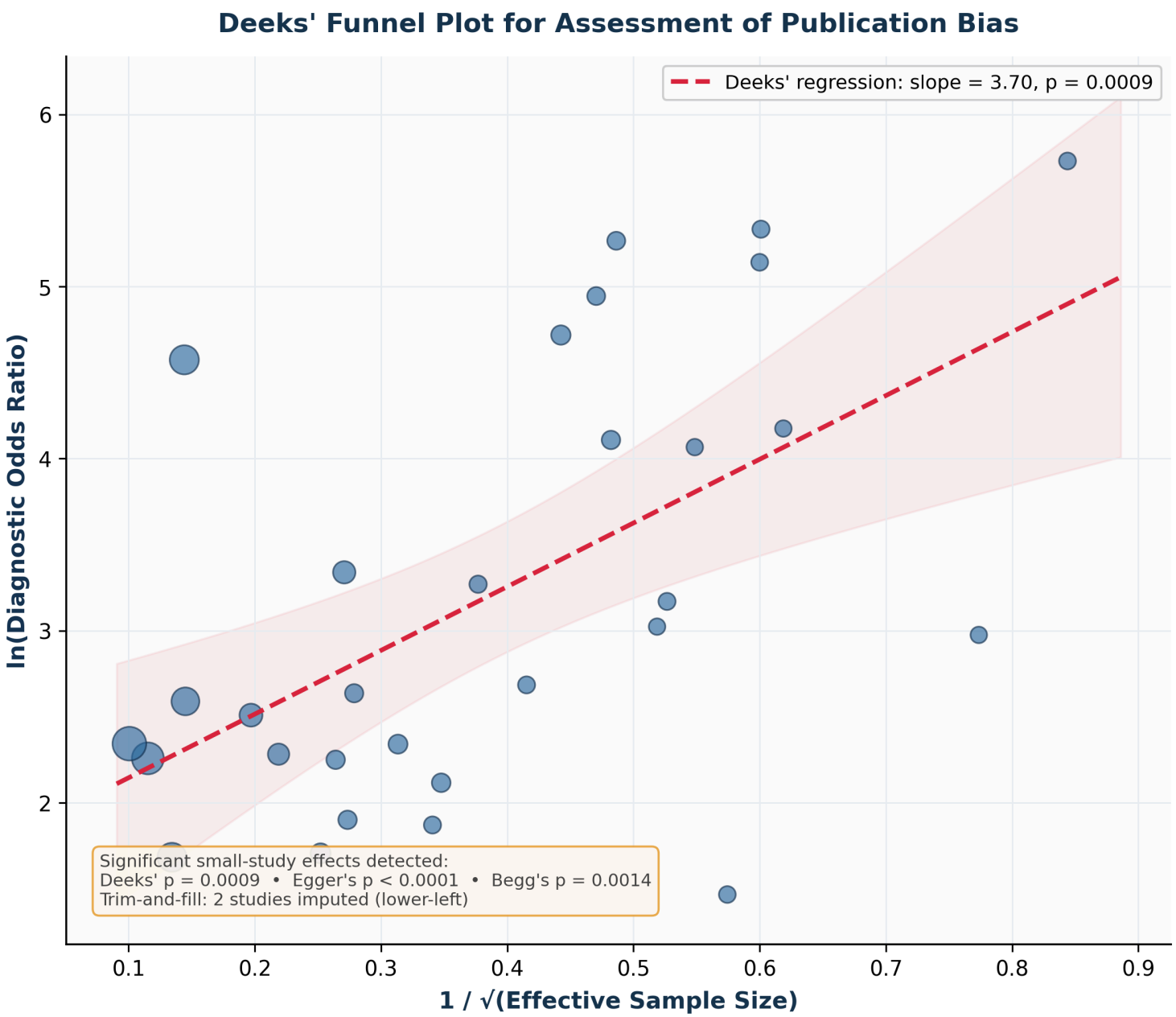


Figure 9

Deeks' funnel plot for publication bias. Log diagnostic odds ratio plotted against the inverse root of effective sample size, with the Deeks' regression line and 95% band. Significant asymmetry (slope 3.70, p = 0.0009) indicates small-study effects, corroborated by Egger's (p < 0.0001) and Begg's (p = 0.0014) tests.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- PSMAMetaAnalysisSupplementaryTablesFinal.docx