Diagnostic Accuracy of 18F-NaF PET/CT versus 99mTc Bone Scintigraphy for Detection of Skeletal Metastases: A Critical Review of Comparative Evidence

Tosif Quazi[1], Haroon Rashid[1], Aamir Majeed Shah[1], Elena Khalid[2]

[1] Department of Radiodiagnosis and Imaging, Sher-i-Kashmir Institute of Medical Sciences, Srinagar, India.
[2] Weston General Hospital, Grange Road, Uphill, BS23 4TQ, United Kingdom.

Correspondence: Haroon Rashid · haroonrashid3805@gmail.com

Original publication: 29 November 2025

Publisher: SJORANM GmbH, Rosenweg 3, CH-6340 Baar, Switzerland.

Abstract

Objective: To reassess the comparative diagnostic evidence for 18F-sodium fluoride (NaF) PET/CT and 99mTc bone scintigraphy following inconsistencies in the original quantitative synthesis.

Methods: This critical review audited primary reports and accessible supplements following PubMed and Europe PMC searches covering publications through 31 March 2025. Comparisons were separated by scintigraphic protocol. Verified patient-based counts yielded study-specific sensitivity and specificity with exact 95% binomial confidence intervals. Incomplete source verification prevented a definitive replacement meta-analysis; this is not a completed systematic review.

Results: Five selected paired comparisons with planar scintigraphy comprised 266 participants. Under the stated study-specific interpretation rules, observed NaF PET/CT sensitivity ranged from 88.9% to 100%, versus 52.3% to 78.9% for planar scintigraphy. Specificity ranged from 90.0% to 100% and from 57.1% to 100%, respectively. These descriptive ranges are not pooled estimates. Additional studies differed in comparator protocol, analysis population or data availability.

Conclusion: NaF PET/CT showed higher observed sensitivity in the verified planar comparisons, but generalization across cancers and scintigraphic protocols is not established. The original pooled claims are not retained. This review does not establish improved patient outcomes or cost-effectiveness.

Keywords: sodium fluoride; PET/CT; bone scintigraphy; skeletal metastases; diagnostic accuracy; critical review

Open Access · CC BY 4.0 · © 2025 The Author(s) · Published by SJORANM GmbH, Switzerland

1. Introduction

Detection of skeletal metastases can influence cancer staging and treatment planning. However, “bone scintigraphy” describes several acquisition and interpretation strategies. A planar scan, planar imaging supplemented by regional SPECT, whole-body SPECT, and integrated SPECT/CT are not interchangeable comparators. Likewise, acquisition on a PET/CT scanner does not establish that the evaluated index reading incorporated CT. These distinctions affect the question answered by each comparison [1,2,6,11].

The original article contained incompatible study totals and quantitative claims that could not be reproduced from its cited sources. This correction replaces those claims with an explicitly bounded critical review. Its purpose is to identify what the accessible comparative evidence supports and to show which data can be independently recalculated. It does not present the available subset as an exhaustive account of all eligible published studies.

2. Methods

Scope and retrieval. The source audit began with the original article and its candidate references. A retrospective search executed on 12 September 2026 covered publications dated through 31 March 2025, retaining the original review’s historical cutoff. The successful searches retrieved 230 PubMed records and 180 Europe PMC records; 234 unique records remained after deduplication. The two sources overlap substantially. Exact queries, exports and retrieval limitations are preserved in the accompanying search record. Reference checking and targeted searches were used to locate original reports and public author or institutional copies.

These retrieval counts are not a completed PRISMA study-selection flow. Full eligibility adjudication and full-text retrieval were not completed for every potentially relevant record. Searches in Embase, Scopus, Web of Science or the Cochrane Library are not claimed. No prospective protocol registration or independent duplicate screening/extraction is claimed. The correction is therefore labelled a critical review, rather than a completed systematic review or meta-analysis [16].

Extraction. The core presentation concerns paired, patient-based comparisons of integrated NaF PET/CT and 99mTc scintigraphy. Counts were accepted from explicit primary tables, exact numerators and denominators, or individual classifications in primary supplementary material. Integer counts were not inferred solely from rounded sensitivity or specificity. PET-only readings, lesion-based outcomes, independent cohorts and unmatched subgroups were kept separate. Unavailable or inconsistent information was recorded rather than imputed.

Analysis. Sensitivity was TP/(TP+FN) and specificity TN/(TN+FP). Two-sided 95% Clopper-Pearson confidence intervals were calculated using the beta distribution. Zero cells received no continuity correction. The tables state each threshold rule; a uniform clinical threshold is not assumed. No overall diagnostic odds ratio, summary ROC area, pooled difference, publication-bias test or meta-regression is reported.

3. Verified patient-based planar comparisons

Table 1 presents five source-checked comparisons and their common patient denominators [1,3-6]. The total of 266 belongs only to this displayed set. It is not the sample size of a completed systematic review. Both tests have the same disease-positive and disease-negative denominators within each displayed comparison.

Table 1. Study-specific diagnostic accuracy. Counts are TP, FP, TN and FN; percentages are followed by exact 95% confidence intervals.

Study / n	Test	TP	FP	TN	FN	Sensitivity %	Specificity %
Even-Sapir 2006 [1] n=44	NaF PET/CT	23	0	21	0	100.0 (85.2-100.0)	100.0 (83.9-100.0)
Even-Sapir 2006 [1] n=44	Planar BS	13	9	12	10	56.5 (34.5-76.8)	57.1 (34.0-78.2)
Ota 2014 [3] n=10	NaF PET/CT	8	0	2	0	100.0 (63.1-100.0)	100.0 (15.8-100.0)
Ota 2014 [3] n=10	Planar BS	6	0	2	2	75.0 (34.9-96.8)	100.0 (15.8-100.0)
Fonager 2017 [4] n=37	NaF PET/CT	24	1	9	3	88.9 (70.8-97.6)	90.0 (55.5-99.7)
Fonager 2017 [4] n=37	Planar BS	21	1	9	6	77.8 (57.7-91.4)	90.0 (55.5-99.7)
Wang 2022 [5] n=58	NaF PET/CT	19	3	36	0	100.0 (82.4-100.0)	92.3 (79.1-98.4)
Wang 2022 [5] n=58	Planar BS	15	10	29	4	78.9 (54.4-93.9)	74.4 (57.9-87.0)
Usmani 2022 [6] n=117	NaF PET/CT	42	6	67	2	95.5 (84.5-99.4)	91.8 (83.0-96.9)
Usmani 2022 [6] n=117	Planar BS	23	31	42	21	52.3 (36.7-67.5)	57.5 (45.4-69.0)

BS, bone scintigraphy; TP, true positive; FP, false positive; TN, true negative; FN, false negative. The five studies represent different cancers and diagnostic settings. Estimates must not be read as interchangeable measures of one common population.

Even-Sapir: Table 2 counts with equivocal results positive were used. Its disease-negative cells sum to 21, although the heading prints 11; the text supports 21 nonmetastatic participants. The paper's abstract also gives a planar sensitivity that differs from Table 2. This unresolved reporting discrepancy is disclosed rather than concealed [1].

Ota: Eleven patients had PET/CT and SPECT, but only ten had planar imaging. Patient 11 was metastatic and PET/CT-positive; subtracting this patient gives the matched PET/CT table shown here. Specificity rests on only two nonmetastatic participants [3].

Fonager: Author-reported dichotomous totals were retained. They differ from classifying every equivocal graded finding as positive. **Wang:** The published binary table was used. **Usmani:** Equivocal findings were positive; individual supplementary classifications independently support the displayed planar and PET/CT counts [4-6].

Observed sensitivities in this selected set favour NaF PET/CT, but several confidence intervals are wide. In particular, an observed specificity of 100% based on two nonmetastatic patients is compatible with much lower underlying specificity. No paired significance test was inferred from these marginal tables.

4. Precision and the handling of equivocal findings

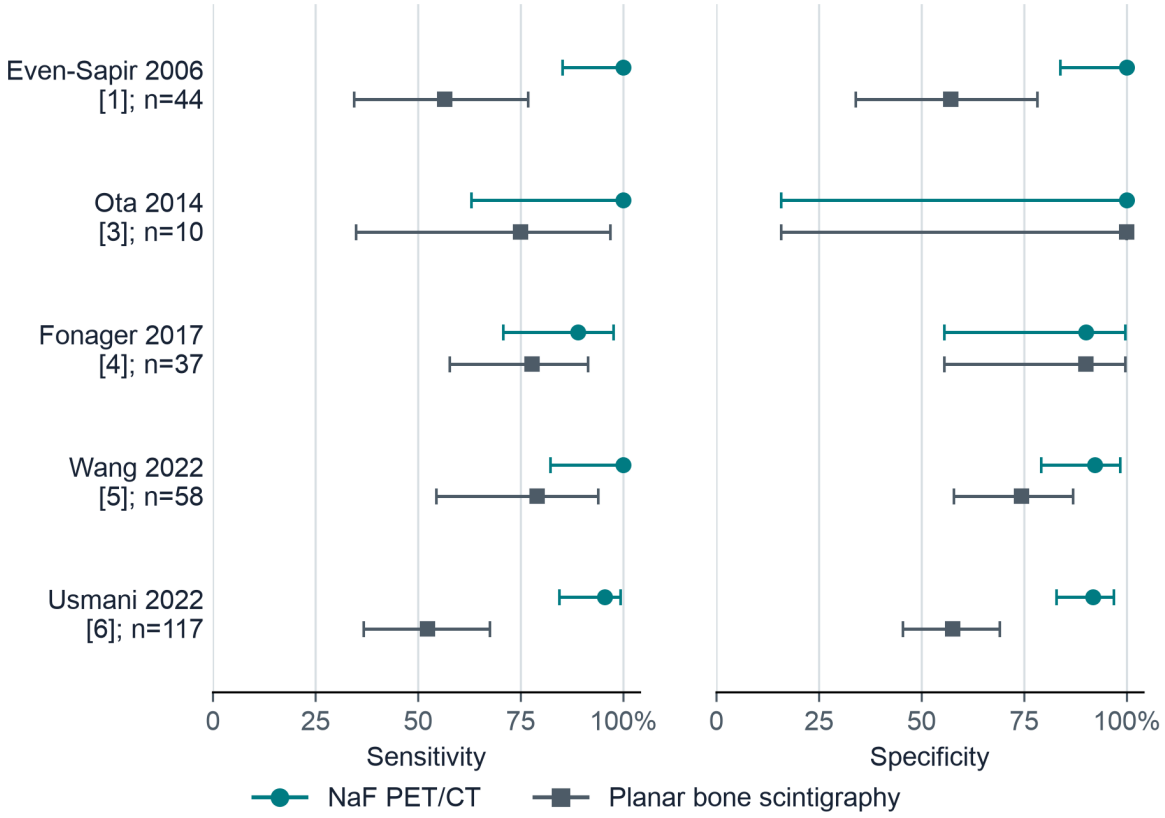


Figure 1. Patient-based sensitivity and specificity for the five planar comparisons in Table 1. Points show observed proportions; bars show exact 95% binomial confidence intervals. No pooled diamond or summary estimate is included. The two tests are plotted within the same study, but paired covariance is not inferred from the marginal counts.

An equivocal result is clinically different from a confidently positive or negative result. Its analytical classification changes sensitivity and specificity and must be stated before estimates are compared. Table 1 uses equivocal-positive interpretations for Even-Sapir and Usmani and the reported binary outcomes for the other sources; it does not claim a harmonized threshold [1,3-6].

In Usmani's 117 patients, treating equivocal findings as negative leaves NaF sensitivity at 95.5% but changes specificity from 91.8% to 95.9%. Planar sensitivity stays at 52.3%, while specificity changes from 57.5% to 71.2%. These changes arise from the primary individual classifications, not from rounding published percentages [6].

For Even-Sapir, using equivocal-negative Table 2 counts changes NaF PET/CT sensitivity from 100% to 87.0% and planar sensitivity from 56.5% to 34.8%. The interpretation rule is therefore a material part of the result. The reporting discrepancy between that table and the abstract limits any unqualified use of a single estimate [1].

5. Tomographic comparisons answer different questions

Table 2 preserves several available tomographic or combined-protocol comparisons. It deliberately does not pool them or add their participants to those in Table 1. Ota and Usmani reappear because the same patients underwent additional comparator protocols [2,3,6,10].

Table 2. Source-checked counts for combined or tomographic protocols. These rows are descriptive; acquisition, interpretation and verification differ.

Study / n	Test	TP	FP	TN	FN	Sensitivity %	Specificity %
Withofs 2011 [2] n=33	NaF PET/CT	20	1	12	0	100.0 (83.2-100.0)	92.3 (64.0-99.8)
Withofs 2011 [2] n=33	Planar + regional SPECT	19	4	9	1	95.0 (75.1-99.9)	69.2 (38.6-90.9)
Ota 2014 [3] n=11	NaF PET/CT	9	0	2	0	100.0 (66.4-100.0)	100.0 (15.8-100.0)
Ota 2014 [3] n=11	SPECT	9	0	2	0	100.0 (66.4-100.0)	100.0 (15.8-100.0)
Usmani 2022 [6] n=117	NaF PET/CT	42	6	67	2	95.5 (84.5-99.4)	91.8 (83.0-96.9)
Usmani 2022 [6] n=117	SPECT/CT	29	11	62	15	65.9 (50.1-79.5)	84.9 (74.6-92.2)
MITNEC-A1 2022 [10] n=261	NaF PET/CT	86	18	134	23	78.9 (70.0-86.1)	88.2 (81.9-92.8)
MITNEC-A1 2022 [10] n=261	Bone scintigraphy with SPECT	69	19	133	40	63.3 (53.5-72.3)	87.5 (81.2-92.3)

Withofs: The comparator incorporated planar imaging and single-field-of-view SPECT. The 33-person patient analysis excludes one of 34 enrolled participants. It must not be relabelled a pure planar study. **Ota:** The eleven-person SPECT comparison is distinct from the ten-person planar comparison. **Usmani:** Equivocal-positive PET/CT and SPECT/CT classifications are shown [2,3,6].

MITNEC-A1: Counts were checked in the original table reproduced in a public EANM presentation (PDF page 10), alongside the primary abstract. Of 288 enrolled participants, 261 contributed to the reported main analysis, including 109 with metastases. The complete Lancet methods and alternative definitions were not obtained for this audit; the comparator is labelled as reported, without assuming an unverified CT protocol [10].

MITNEC-A1 reported a sensitivity difference of 15.6 percentage points in favour of NaF PET/CT and no statistically significant specificity difference under its primary definition. This is an individual trial result, not an overall comparative effect [10]. The other displayed comparisons likewise cannot establish a universal superiority over every SPECT or SPECT/CT implementation.

6. Studies and data requiring separate treatment

The records below explain why a source being clinically relevant does not make it suitable for the same numerical synthesis. An unavailable full text is a retrieval limitation, not evidence that the study is invalid or that its findings are negative.

Table 3. Material distinctions and limitations identified in additional primary reports.

Report	Verified scope	Treatment in this correction
Jambor / SKELETA 2016 [7]	53 patients; multiple modalities; complete report retrieved.	Narrative. Rounded patient-level estimates and threshold-dependent reporting do not provide a sufficiently unambiguous complete table for this audit. No integer back-calculation was used.
Löfgren 2017 [8]	117 patients; planar and SPECT/CT versus NaF PET/CT.	Primary abstract verified; full patient tables not obtained. Relevant evidence is acknowledged, but no counts or pooled contribution are assigned.
Dyrberg 2018 [9]	213 patients had NaF PET/CT; 109 also had SPECT/CT.	Full report and supplement retrieved. The available overall NaF table is not the 109-person comparator subset. No matched NaF counts are invented.
Abikhzer 2016 [11]	41 women; NaF reading labelled PET, compared with whole-body SPECT.	Kept separate from integrated PET/CT interpretation. The different 92-patient scintigraphy study cannot supply a NaF row.
Rao 2016 [12]	181 NaF PET/CT patients and 167 SPECT patients.	Independent cohorts; not a paired comparison. The NaF arm alone is not the study's total population.
Schirrmeister 2001 [13]	53 patients; PET without integrated CT.	Table 1 is internally inconsistent: its categories imply 13 diseased patients, while the text reports 12. The earlier reconstructed table is not reused.
Damle 2013 [14]; Chakraborty 2013 [15]	151 patients with breast/lung/prostate cancer; 48 with bladder cancer, respectively.	Primary abstracts establish relevant comparisons. Full source tables were not obtained; no reconstructed counts are substituted from secondary reviews.

These unresolved and differently defined datasets are a reason to retain a bounded descriptive scope. Omitting them silently while presenting a conveniently available subset as the complete meta-analysis would create selection and applicability problems. A later systematic synthesis should retrieve and adjudicate these reports, resolve overlap where relevant, and document exclusions before locking the analysis dataset.

7. Appraisal of validity and applicability

The appraisal follows the domains of patient selection, index test, reference standard, and flow/timing used in QUADAS-2 [17]. It is a structured editorial appraisal of accessible information, not a completed independent duplicate QUADAS-2 assessment. Missing information was not interpreted as evidence of low risk.

Table 4. Main source-supported concerns and their implications.

Report(s)	Observation from accessible material	Implication
Even-Sapir [1]	High-risk prostate population; CT and follow-up inform lesion characterization; internal reporting discrepancies.	Spectrum and reference-standard concerns; the table/abstract inconsistency limits confidence in one numerical estimate.
Withofs [2]	Selected breast/prostate population; targeted MRI or CT confirmation depended on lesion location and findings; one patient excluded.	Selective/differential verification can influence accuracy; the sample does not represent unselected staging.
Fonager [4]	High-risk newly diagnosed prostate cancer; consensus reference assessment includes imaging and follow-up.	Restricted spectrum and potential incorporation bias; graded and binary interpretation rules must remain distinct.
Wang [5]	Retrospective NPC cohort; pathological evidence or imaging follow-up, including PET/CT, informs the final diagnosis.	Potential selection and incorporation bias; performance cannot be generalized across cancer types.
Usmani [6]	BMI >40 kg/m²; blinded index readings; radiological/clinical follow-up of at least six months; incomplete follow-up excluded.	Useful defined population, but limited generalizability; reference blinding and exclusions require attention.
Jambor; Dyrberg [7,9]	Final diagnosis uses consensus including index tests; Dyrberg accepts concordant imaging as final diagnosis.	Index-test incorporation and nonuniform verification may overestimate accuracy.
Ota; MITNEC-A1; Löfgren [3,10,8]	Only primary tables/abstract or a reproduced table available for parts of the audit.	Full domain-level appraisal remains incomplete. Partial access cannot justify a low-risk judgement.

Histology for every skeletal focus is rarely practical, but a composite reference standard is not automatically unbiased. The key questions are whether verification is applied consistently, whether assessors know index-test results, and whether the index test helps define its own correctness. Long follow-up alone does not remove these concerns.

No global quality score or overall certainty grade is assigned. Several studies are small, restricted to a particular cancer or risk group, and affected by sparse disease-negative or disease-positive denominators. Their observed accuracy should be interpreted within those populations.

8. Discussion

The source-checked planar comparisons consistently show higher observed NaF PET/CT sensitivity under the stated interpretation rules. The size and precision of that difference vary, and the clinical implications depend on cancer type, disease spectrum, comparator protocol and reference standard. This supports a carefully defined diagnostic role, not a blanket replacement recommendation [1,3-6].

Tomographic results require particular care. Adding regional SPECT or CT changes the comparator, and analyzing different patient subsets can produce an apparent comparison that is not truly paired. Dyrberg's overall NaF results, for example, must not be paired numerically with the smaller SPECT/CT cohort without the corresponding NaF subset [9].

Why the original meta-analysis is not retained. The prior pooled estimates depended on an unverified and inconsistently defined study set. An exploratory feasibility calculation on five available planar tables reached the correlation boundary in the full bivariate model. This does not prove that all meta-analysis is impossible: simpler binomial hierarchical models can be appropriate for sparse data [18]. However, changing the estimator cannot correct missing relevant studies, unclear test definitions or unresolved source errors. No exploratory pooled estimate is promoted to a final result in this correction.

Clinical conclusions. Diagnostic accuracy, a change in intended management, improved patient outcomes and cost-effectiveness are different outcomes. The present audit does not establish a survival benefit, a reduction in skeletal events, or cost-effectiveness. Decisions about staging pathways require current disease-specific evidence and the local availability and performance of alternative tests.

Limitations. This is a retrospective, AI-assisted critical source audit with incomplete full-text retrieval and no independent duplicate human screening or extraction. Its search used two overlapping public bibliographic sources and a historical publication cutoff. Selected verified comparisons illustrate the recoverable evidence; they are not an exhaustive inclusion set. These limitations prevent this version from being described as a completed replacement systematic review.

9. Conclusion

NaF PET/CT had higher observed sensitivity than planar bone scintigraphy in the five verified paired comparisons. Generalization to all cancers, all scintigraphic protocols or patient outcomes is not established. The correction retains transparent study-level data and removes the original unsupported pooled claims. A definitive comparative meta-analysis requires completion of the source and eligibility work first.

Declarations

Funding and competing interests: The original author declarations report no funding and no conflicts of interest. **Ethics and consent:** This review uses published aggregate information and no identifiable patient data or clinical images; new participant consent is not applicable. **Data availability:** Extraction tables, exact search records and calculation outputs are supplied with the correction package. **AI assistance:** Codex (OpenAI) assisted retrieval, extraction, analysis, drafting and typesetting. Responsibility for verification and approval remains with the authors and editorial team.

References

1. Even-Sapir E, Metser U, Mishani E, et al. The detection of bone metastases in patients with high-risk prostate cancer: 99mTc-MDP Planar bone scintigraphy, single- and multi-field-of-view SPECT, 18F-fluoride PET, and 18F-fluoride PET/CT. J Nucl Med. 2006;47:287-297. https://pubmed.ncbi.nlm.nih.gov/16455635/

2. Withofs N, Grayet B, Tancredi T, et al. 18F-fluoride PET/CT for assessing bone involvement in prostate and breast cancers. Nucl Med Commun. 2011;32:168-176. doi: 10.1097/MNM.0b013e3283412ef5

3. Ota N, Kato K, Iwano S, et al. Comparison of 18F-fluoride PET/CT, 18F-FDG PET/CT and bone scintigraphy (planar and SPECT) in detection of bone metastases of differentiated thyroid cancer: a pilot study. Br J Radiol. 2014;87:20130444. doi: 10.1259/bjr.20130444

4. Fonager RF, Zacho HD, Langkilde NC, et al. Diagnostic test accuracy study of 18F-sodium fluoride PET/CT, 99mTc-labelled diphosphonate SPECT/CT, and planar bone scintigraphy for diagnosis of bone metastases in newly diagnosed, high-risk prostate cancer. Am J Nucl Med Mol Imaging. 2017;7:218-227. https://pubmed.ncbi.nlm.nih.gov/29181269/

5. Wang D, Yang Y, Zeng Z, et al. Comparison of Bone Metastases between 18F-NaF PET/CT, 18F-NaF PET, and Planar 99mTc-MDP Bone Scintigraphy in Patients with Newly Diagnosed Nasopharyngeal Carcinoma. Contrast Media Mol Imaging. 2022;2022:5975338. doi: 10.1155/2022/5975338

6. Usmani S, Ahmed N, Gnanasegaran G, et al. Prospective study of Na[18F]F PET/CT for cancer staging in morbidly obese patients compared with [99mTc]Tc-MDP whole-body planar, SPECT and SPECT/CT. Acta Oncol. 2022;61:1230-1239. doi: 10.1080/0284186X.2022.2101899

7. Jambor I, Kuisma A, Ramadan S, et al. Prospective evaluation of planar bone scintigraphy, SPECT, SPECT/CT, 18F-NaF PET/CT and whole body 1.5T MRI, including DWI, for the detection of bone metastases in high risk breast and prostate cancer patients: SKELETA clinical trial. Acta Oncol. 2016;55:59-67. doi: 10.3109/0284186X.2015.1027411

8. Löfgren J, Mortensen J, Rasmussen SH, et al. A Prospective Study Comparing 99mTc-Hydroxyethylene-Diphosphonate Planar Bone Scintigraphy and Whole-Body SPECT/CT with 18F-Fluoride PET/CT and 18F-Fluoride PET/MRI for Diagnosing Bone Metastases. J Nucl Med. 2017;58:1778-1785. doi: 10.2967/jnumed.116.189183

9. Dyrberg E, Hendel HW, Løgager VB, et al. A prospective study determining and comparing the diagnostic accuracy of fluoride-PET/CT, choline-PET/CT, whole-body bone SPECT/CT and whole-body MRI for the detection of bone metastases in patients with prostate cancer. Eur J Hybrid Imaging. 2018;2:19. doi: 10.1186/s41824-018-0038-y

10. Bénard F, Harsini S, Wilson D, et al. Intra-individual comparison of 18F-sodium fluoride PET-CT and 99mTc bone scintigraphy with SPECT in patients with prostate cancer or breast cancer at high risk for skeletal metastases (MITNEC-A1): a multicentre, phase 3 trial. Lancet Oncol. 2022;23:1499-1507. doi: 10.1016/S1470-2045(22)00642-8

11. Abikhzer G, Srour S, Fried G, et al. Prospective comparison of whole-body bone SPECT and sodium 18F-fluoride PET in the detection of bone metastases from breast cancer. Nucl Med Commun. 2016;37:1160-1168. doi: 10.1097/MNM.0000000000000568

12. Rao L, Zong Z, Chen Z, et al. 18F-Labeled NaF PET-CT in Detection of Bone Metastases in Patients With Preoperative Lung Cancer. Medicine (Baltimore). 2016;95:e3490. doi: 10.1097/MD.0000000000003490

13. Schirrmeister H, Glatting G, Hetzel J, et al. Prospective evaluation of the clinical value of planar bone scans, SPECT, and (18)F-labeled NaF PET in newly diagnosed lung cancer. J Nucl Med. 2001;42:1800-1804. https://pubmed.ncbi.nlm.nih.gov/11752076/

14. Damle NA, Bal C, Bandopadhyaya GP, et al. The role of 18F-fluoride PET-CT in the detection of bone metastases in patients with breast, lung and prostate carcinoma: a comparison with FDG PET/CT and 99mTc-MDP bone scan. Jpn J Radiol. 2013;31:262-269. doi: 10.1007/s11604-013-0179-7

15. Chakraborty D, Bhattacharya A, Mete UK, et al. Comparison of 18F fluoride PET/CT and 99mTc-MDP bone scan in the detection of skeletal metastases in urinary bladder carcinoma. Clin Nucl Med. 2013;38:616-621. doi: 10.1097/RLU.0b013e31828da5cc

16. McInnes MDF, Moher D, Thombs BD, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. JAMA. 2018;319:388-396. doi: 10.1001/jama.2017.19163

17. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155:529-536. doi: 10.7326/0003-4819-155-8-201110180-00009

18. Takwoingi Y, Guo B, Riley RD, Deeks JJ. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. Stat Methods Med Res. 2017;26:1896-1911. doi: 10.1177/0962280215592269