Journal of Nuclear Medicine, published on September 3, 2019 as doi:10.2967/jnumed.119.232686

Observer agreement and accuracy of 18F-sodium-fluoride PET/CT in the diagnosis of bone metastases in prostate cancer

Helle D. Zacho[1,2], Randi F Fonager[1], Julie B Nielsen[1], Christian Haarmark[3], Helle W Hendel[3], Martin B Johansen[4] ,Jesper C. Mortensen[5], Lars J. Petersen[1,2]

[1] Dept. of Nuclear Medicine, Clinical Cancer Research Center, Aalborg University Hospital, [2]Dept. of Clinical Medicine, Aalborg University, Denmark, [3]Dept. of Nuclear Medicine, Herlev and Gentofte Hospital, Denmark, [4]Unit of Clinical Biostatistics, Aalborg University Hospital, [5]Dept. of and Nuclear Medicine, Regional Hospital West Jutland, Denmark

The data were presented in part at the 2017 Annual Meeting in the European Association of Nuclear Medicine (electronic abstract EP-0712)

Word count: 5000

Corresponding author

Helle D Zacho

Dept. of Nuclear Medicine

Aalborg University Hospital

Hobrovej 18-22

Postboks 365

DK-9100 Aalborg

Denmark

Tel (+45) 97 66 55 00

Fax (+45) 97 66 55 01

E-mail: h.zacho@rn.dk

Financial support: The study was supported by an unrestricted grant from the Danielsen Foundation, the Obel Family Foundation, the Danish Medical Research Grant/The Højmosegård Grant, and the Heinrich Kopps Grant.

Running head: Interobserver agreement in NaF PET/CT

ABSTRACT

Aim: To evaluate the interobserver agreement in 18F-sodium fluoride (NaF) PET/CT for the detection of bone metastases in patients with prostate cancer (PCa).

Materials and Methods: NaF PET/CT scans were retrieved from all patients who participated in four recent prospective trials. Two experienced observers independently evaluated the NaF PET/CTs on a patient level using a three-category scale (no bone metastases [M0], equivocal for bone metastases, and bone metastases present [M1]) and on a dichotomous scale (M0/M1). In patients with no more than 10 lesions, the location and number of lesions were recorded. On a patient level, the diagnostic performance was calculated using a sensitivity analysis, where equivocal lesions were handled as M0 as well as M1.

Results: NaF PET/CTs from 219 patients with PCa were included, of whom 129 patients were scanned for primary staging, 67 patients for biochemical recurrence, and 23 patients for metastatic castration-resistant PCa. Agreement between the observers was almost perfect on a patient level (three-category scale unweighted κ: 0.83 ± 0.05, linear weighted κ: 0.90 ± 0.06; dichotomous scale κ: 0.91 ± 0.07). On a lesion level (dichotomous scale), the observers agreed on the number as well as the location of bone metastases in 205 (93.6%) patients. In the remaining 14 patients, the readers disagreed on the number of lesions in 13 patients and the location of bone metastases in one patient. A final diagnosis of bone metastases was made in 211 of 219 patients. The sensitivity ranged from 0.86-0.92, specificity from 0.83-0.97, positive predictive value from 0.70-0.93, and negative predictive value from 0.94-0.96.

Conclusion: The interobserver agreement of NaF PET/CT for the detection of bone metastases in patients with PCa was very high among trained observers, both on a patient level and on a lesion level.

Moreover, the diagnostic performance of NaF PET/CT was satisfactory, rendering NaF PET/CT a robust tool in the diagnostic armamentarium.

Keywords: 18F-sodium-fluoride PET/CT, prostate cancer, interobserver agreement, kappa-value, bone metastases.

INTRODUCTION

Prostate cancer (PCa), the most common cancer in European men (*1*), has a predilection for metastasizing to the bones. A correct diagnosis of bone metastases is fundamental to determine the correct treatment for an individual patient. The European Association of Urology and the National Comprehensive Cancer Network recommend bone scintigraphy for assessment of bone metastases (*2,3*).

With increasing access to PET/CT, several centers have replaced bone scintigraphy with 18F-sodium fluoride (NaF) PET/CT for the assessment of bone metastases. NaF PET/CT exhibits improved pharmacokinetic properties and a favorable target-background ratio compared to methylene diphosphonate, which is frequently used for bone scintigraphy (*4*). Additionally, several studies imply improved diagnostic accuracy by NaF PET/CT compared to bone scintigraphy (*5-7*).

Despite favorable properties in NaF PET/CT, imaging evaluation remains a matter of subjectivity and has been labeled the weakest aspect of imaging (*8,9*). We have previously shown excellent interobserver agreement among trained nuclear medicine physicians for the assessment of bone metastases by planar bone scintigraphy (*10*), but to the best of our knowledge, no such studies have been conducted using NaF PET/CT.

The aim of the present exploratory study was primarily to evaluate the interobserver agreement for NaF PET/CT findings among experienced observers in various settings of prostate cancer and secondly to assess diagnostic accuracy of bone metastasis in patients with PCa.

MATERIALS AND METHODS

Patients

We included patients with PCa who had participated in four prospective studies at our department (*6,11-13*). All patients had undergone NaF PET/CT as part of the study-related procedures. If more than one NaF PET/CT scan had been conducted, the first NaF PET/CT was included in the present analysis. The patients represent a wide range of disease stages from newly diagnosed PCa and biochemical recurrence after previous curative intent treatment to metastatic castration-resistant PCa.

NaF PET/CT

NaF PET/CT was conducted in accordance with the guidelines from the Society of Nuclear Medicine (*14*) and the European Association of Nuclear Medicine (*15*) as previously described (*6*). In short, image acquisition was performed in 3D mode from the vertex to the mid-thigh 30 min after intravenous injection of approximately 200 MBq of 18F-sodium fluoride. The images were reconstructed using attenuation correction, an ordered subset expectation maximization algorithm. Immediately after the PET acquisition, low-dose CT was conducted. Low-dose CT was used for attenuation correction and anatomical coregistration.

Blinded evaluation of NaF PET/CT findings

Each patient's set of NaF PET/CT images was independently interpreted by two board-certified nuclear medicine physicians. Both readers were very experienced (with experience of more than 2000 NaF PET/CT evaluations at the start of the study). The observers were blinded to all clinical information except for the PCa diagnosis.

The NaF PET/CT findings were categorized on a patient level into one of three categories: 1) no bone metastases (M0), 2) equivocal, and 3) bone metastases (M1). In addition, the observers had to classify the NaF PET/CT findings dichotomously as bone metastases being present or not present. For patients with no more than ten lesions, the observers marked metastatic and equivocal lesions on a schematic drawing of a skeleton to evaluate whether they were considering the same lesions metastatic in each patient. Patients with a metastatic superscan were included in the category with 10+ bone metastases.

Clinical data and final diagnosis of metastatic status

Clinical data were retrieved from the case report forms from the prospective studies. Patients were categorized with biochemical recurrence and castration-resistant PCa based on the criteria from the European Association of Urology (*2,16*). In the majority (211 of 219 patients, 96%) of the patients, a final diagnosis for the presence or absence of bone metastases was available from the initial study in which the patient participated (*6,11-13*). In short, in every study a final diagnosis on a patient level was achieved by combing all clinical data, biochemical data and imaging conducted prior to inclusion in the study and during follow up. This was done by a group of experienced nuclear medicine physicians, a radiologist and an urologist; In two studies, the patients underwent systematic imaging follow up (*6,11*). In one study, the patients underwent additional MR, diffusion weighted MRI and at least two years of follow-up was available for all patients (*12*). Finally, in one study, all patients underwent radical prostatectomy. All patients with post-operative PSA values below 0.1 ng/mL for at least two years after surgery was classified without bone metastasis at staging (*13*).

Statistics

The agreement between the two observers was assessed on both the original rating in one of the three categories and the dichotomous classification. For the three-category scale, we used Cohen's Kappa (κ) calculated as unweighted and with a linear weighting, treating the equivocal option as 0.5. The κ was categorized according to the suggestion by Landis and Koch (*17*). For the dichotomous scale, we calculated κ as well as the positive and negative agreement (p_{pos} and p_{neg}) defined as twice the number of cases for which the two observers agreed on a positive (negative) rating divided by the sum of the number of cases for which each observer gave a positive (negative) rating. Because the p_{pos} and p_{neg} are not simple proportions of independent observations, we calculated percentile-based confidence intervals from a bootstrapping procedure with 1000 replications for these measures. All other confidence intervals are analytical.

Furthermore, the observer interpretation of the NaF PET/CT was evaluated against the final diagnosis (used as the reference) by calculating the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for each observer as well as for the consensus rating by the two observers. For the three-category scale, we calculated diagnostic accuracy using sensitivity analysis where equivocal findings were considered either M0 or M1. Statistical analyses were performed using STATA®15 (StataCorp LP, College Station, TX, USA).

Approvals

This study complied with the Helsinki II Declaration. All patients provided signed informed consent to participate. The studies were approved by the ethical committee (N-20130068, N-20140042, N-20140057 and N-20140080) and by the Danish Data Protection Agency.

RESULTS

Patients

A total of 219 patients were included in the analysis of interobserver agreement. The patients had participated in four recent prospective studies (*6,11-13*), and NaF PET/CT was performed as a part of the study-related procedures. NaF-PET/CT was performed at primary staging in 129 patients, at the time of biochemical recurrence after previous treatment with curative intent in 67 patients, and finally, NaF PET/CT was performed in 23 patients with metastatic castration-resistant PCa (Table 1).

Observer agreement on a patient level

Using the three-category scale, crude agreement between the two observers was seen in 200/219 (91%) of the patients, corresponding to an unweighted κ value of 0.83 ± 0.05 (linear weighted κ was 0.90 ± 0.06). The observers agreed that 131 patients had no bone metastases, 61 patients had bone metastases and the NaF PET/CT findings were equivocal in eight patients. The observers disagreed in 19 patients (supplemental Table 1); in one patient, the observers had a two-category difference (no bone metastases vs. bone metastases), whereas in the remaining 18 patients, the difference in categorization was equivocal vs. no bone metastases or equivocal vs. bone metastases.

Using the dichotomous scale, the crude agreement increased. The observers agreed on the M0/M1 classification in 211/219 (96%) of the patients (κ: 0.91 ± 0.07). The p_{pos} (M1) was 0.94, and the p_{neg} (M0) was 0.97.

In addition, the agreement was assessed for each disease stage (Table 2). The best observer agreement was found with metastatic castration-resistant PCa (κ: 1.0) whereas most disagreement was observed in biochemical recurrence (unweighted κ: 0.70, linear weighted κ: 0.79). Overall, disagreement was seen in patients with NaF uptake without corresponding changes on the low-dose CT.

Observer agreement on a lesion level (dichotomous scale)

In patients with bone metastases according to at least one of the observers (n = 72), the number of lesions (from 1 to 10) and lesion location were compared between the observers. Thirty-nine patients had more than ten bone metastases as determined by both observers.

Of the remaining 33 patients, complete agreement regarding both the number and the location of the lesions was observed in 19 patients (Fig. 1). A difference in the number of metastatic lesions was encountered in 13 patients (Fig. 2; Supplemental Fig. 1), of whom one patient was categorized with metastatic superscan by one observer, who also indicated that similar changes may be caused by metabolic bone disease, whereas the other observer indicated benign metabolic bone disease. Finally, the observers indicated different locations of a single lesion out of a total of six lesions considered to be bone metastases in one patient.

Diagnostic characteristics of 18F-NaF PET/CT

In the majority of the patients (211 of 219, 96%), a final diagnosis of bone metastases present or absent was established based on clinical and imaging follow up (*6,11-13*). For both observers and for the consensus evaluation, the sensitivity, specificity, PPV and NPV were calculated for the

dichotomous evaluation and for the three-category evaluation with equivocal results analyzed as negative for bone metastases and then positive for bone metastases (Table 3). In general, the sensitivity ranged from 0.86 to 0.92, and the specificity ranged from 0.83 to 0.97.

DISCUSSION

To the best of our knowledge, the present study is the first to evaluate the interobserver agreement in NaF PET/CT for the detection of bone metastases. Two experienced observers blindly evaluated the NaF PET/CT images for the presence or absence of bone metastases in a large group of patients with PCa and found almost perfect agreement in their assessment of bone metastases. The agreement was investigated across the different stages of PCa with substantial to almost perfect results.

Method evaluation comprised test-retest as well as observer assessments, and previous studies have primarily focused on test-retest evaluation in NaF PET/CT, particularly for quantitative parameters such as standardized uptake values (SUV) (*18-20*), investigated methods for quantification of tumor burden (*20*) or the inter- and intraobserver variation in the assessment of tumor burden (*21*). Our findings of almost perfect agreement are similar to the previously reported interobserver agreement in planar whole-body bone scintigraphy (*10*), which found a κ value of 0.87 for a dichotomous evaluation of bone scintigraphy. In the study of bone scintigraphy, a four-category scale for evaluation was available; this might explain the difference observed between the linear weighted κ values, which was 0.72 for bone scintigraphy versus 0.90 for NaF PET/CT. A similar study using the tracer 68Ga-prostate specific membrane antigen (PSMA) in PET/CT found almost perfect agreement among experienced readers for the presence of bone metastases with κ values of 0.84 in patients with biochemical recurrence and Kappa-values of 0.87 for bone metastases at the time of staging (*22*). Additionally, a large study evaluating PSMA PET/CT in more than 600 patients with biochemical

recurrence reported κ values of 0.78 for the assessment of bone metastases, analogous to the present findings (*23*).

In the present study, a head-to-head comparison of the number and localization of lesions in patients with up to ten bone lesions was conducted. In the vast majority of patients, the observers agreed on both the number and the location of bone metastases. In only a minority of patients (n=13), a difference in the number of metastatic lesions was determined between the two observers, and in only one patient, the observers considered different lesions metastatic. This is an important finding in the STAMPEDE era (Systemic Therapy in Advancing or Metastatic Prostate Cancer: Evaluation of Drug Efficacy), where the number and location of bone metastases is crucial to assign upfront chemotherapy or abiraterone appropriately (*24,25*), or to offer radiotherapy in men with metastatic PCa who have a low metastatic burden (*26*).

Across the range of disease stages, the observer evaluation was compared to a final diagnosis of bone metastases at the patient level. The sensitivity ranged from 0.86 to 0.92, and the specificity was 0.83-0.97. The sensitivity reported in the present study was numerically larger than that reported in a recent prospective study by Löfgren et al (*7*), whereas the specificity was comparable. Löfgren et al used a three-category scale for evaluation, whereas other studies have applied a dichotomous scale for the evaluation (*27,28*) with results comparable to our dichotomous evaluation. A review of NaF PET/CT by Wondergem et al showed a sensitivity of 89%, specificity of 91% for the detection of bone metastases on a patient level (*29*); these number are in line with our data. However, a few studies have reported moderate specificity of NaF PET/CT due to uptake in benign degenerative

and inflammatory lesions. Furthermore, the present diagnostic accuracy for NaF PET/CT was comparable to similar studies using PSMA PET/CT for bone metastases (*23,30*).

The use of a three-category scale provides the opportunity for the equivocal option and may thus resemble "every day clinical practice" more than the dichotomous option does. We found that the proportion of equivocal findings by consensus was considerably lower in the present study of NaF PET/CT than that observed in a similar study using planar bone scintigraphy, in which approximately one in four patients had equivocal findings (*31*). The proportion of equivocal findings in NaF PET/CT resembles previously published findings in studies that added SPECT/CT to planar bone scintigraphy (*6,32,33*) and the reported proportion of equivocal findings in previous diagnostic test accuracy studies in NaF PET/CT (*5,7*). In contrast, the proportion of equivocal NaF PET/CT in the present study was much lower than that stated by the US National PET Registry study, which reported equivocal NaF PET/CT in 15% of the patients (*34*).

To assure the independence and blinding of the observers, they were not involved in recruiting or managing patients in any way. The observers had extensive experience with NaF PET/CT from a high-volume center with many years of experience. The high level of agreement might in part be due to their vast experience and the experience level may render the results less generalizable for readers with less experience and for centers with low volume of NaF PET/CT-scans. To the best of our knowledge, no studies have investigated the impact of reader experience regarding observer agreement or accuracy with NaF PET/CT. The number of NaF PET/CT scans to read to reach the assessment of experts remains unknown, in PSMA-PET/CT this number is considered to be around 300

(*22*). Whether the number could be lower for NaF PET/CT if the observer has extensive experience in bone scintigraphy with SPECT/CT remains undetermined.

According to the US National PET registry study, NaF PET/CT changed patient management in approximately 40% of the patients, which was reduced to 12-16% in patients with plans for other imaging modalities (*34*). In a study by Gauthé et al, NaF PET/CT changed patient management in 7% of patients at initial staging (*35*). However, the improved diagnostic properties of NaF PET/CT or the NaF PET/CT-induced change in patient management have not yet shown improvement in patient-related outcomes. We recently showed that in patients with a negative bone scintigraphy who underwent radical prostatectomy, NaF PET/CT did not add prognostic value in terms of the outcome after radical prostatectomy (*13*). There is a lack of studies showing that NaF PET/CT is beneficial in terms of improved patient-related outcome, which might be one of the reasons why NaF PET/CT is not recommended by the international guidelines for assessment of bone metastases.

In conclusion, the present study demonstrated almost perfect agreement between two observers in using NaF PET/CT images to evaluate bone metastases in PCa. Likewise, the NaF PET/CT agreement on a lesion level was substantial, and the diagnostic accuracy was satisfactory, rendering NaF PET/CT a robust tool for the assessment of bone metastases in PCa. Future studies in NaF PET/CT should focus on the patient-related outcome to evaluate whether the advantageous properties of NaF PET/CT are reflected in patient-related outcomes.

DISCLOSURE

The authors have no potential conflicts of interest relevant to this study.

ACKNOWLEDGMENTS

Study N-20140042 was supported by an unrestricted grant from the Danielsen Foundation.

Study N-20130068, N-20140057 and N-20140080 were supported by unrestricted grants from the Obel Family Foundation.

Study N-20130068, N-20140057 were supported by The Danish Medical Research Grant (The Højmosegård Grant) and the Heinrich Kopps Grant.

Key Points

Questions: What is the interobserver agreement in ^{18}F-sodium fluoride PET/CT (NaF PET/CT) for the detection of bone metastases in patients with prostate cancer (PCa)?

Pertinent findings: Two experienced observers independently evaluated NaF PET/CT from four prospective trials using NaF PET/CT for the detection of bone metastases in 219 patients with PCa across primary staging, biochemical recurrence and metastatic castration resistant PCa. Excellent agreement was seen both on a patient level and on a lesion level for the detection of bone metastases. Furthermore, satisfactory diagnostic accuracy was seen when the findings were compared to a final diagnosis.

Implications for patient care: NaF PET/CT has excellent interobserver agreement and is robust tool for the detection of bone metastases in patients with prostate cancer.

Reference List

1. Taitt HE. Global trends and prostate cancer: a review of incidence, detection, and mortality as influenced by race, ethnicity, and geographic location. *Am J Mens Health.* 2018;12:1807-1823.

2. Mottet N, Bellmunt J, Bolla M, et al. EAU-ESTRO-SIOG Guidelines on prostate cancer. Part 1: screening, diagnosis, and local treatment with curative intent. *Eur Urol.* 2017;71:618-629.

3. Mohler JL, Antonarakis ES, Armstrong AJ, et al. Prostate cancer, version 2.2019, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw.* 2019;17:479-505.

4. Czernin J, Satyamurthy N, Schiepers C. Molecular mechanisms of bone 18F-NaF deposition. *J Nucl Med.* 2010;51:1826-1829.

5. Even-Sapir E, Metser U, Mishani E, Lievshitz G, Lerman H, Leibovitch I. 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.

6. Fonager RF, Zacho HD, Langkilde NC, et al. Diagnostic test accuracy study of (18)F-sodium fluoride PET/CT, (99m)Tc-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.

7. Löfgren J, Mortensen J, Rasmussen SH, et al. A prospective study comparing (99m)Tc-hydroxyethylene-diphosphonate planar bone scintigraphy and whole-body SPECT/CT with (18)F-fluoride PET/CT and (18)F-fluoride PET/MRI for diagnosing bone metastases. *J Nucl Med.* 2017;58:1778-1785.

8. Brealey S, Westwood M. Are you reading what we are reading? The effect of who interprets medical images on estimates of diagnostic test accuracy in systematic reviews. *Br J Radiol.* 2007;80:674-677.

9. Robinson PJ. Radiology's Achilles' heel: error and variation in the interpretation of the Rontgen image. *Br J Radiol.* 1997;70:1085-1098.

10. Zacho HD, Manresa JA, Mortensen JC, Bertelsen H, Petersen LJ. Observer agreement and accuracy in the evaluation of bone scans in newly diagnosed prostate cancer. *Nucl Med Commun.* 2015;36:445-451.

11. Fonager RF, Zacho HD, Langkilde NC, et al. Prospective comparative study of (18)F-sodium fluoride PET/CT and planar bone scintigraphy for treatment response assessment of bone metastases in patients with prostate cancer. *Acta Oncol.* 2018;57:1063-1069.

12. Zacho HD, Nielsen JB, Afshar-Oromieh A, et al. Prospective comparison of (68)Ga-PSMA PET/CT, (18)F-sodium fluoride PET/CT and diffusion weighted-MRI at for the detection of bone metastases in biochemically recurrent prostate cancer. *Eur J Nucl Med Mol Imaging.* 2018;45:1884-1897.

13. Zacho H, Jochumsen MR, Langkilde NC, et al. No added value of (18)F-sodium fluoride PET/CT for the detection of bone metastases in patients with newly diagnosed prostate cancer with normal bone scintigraphy. *J Nucl Med.* May 30, 2019 [Epub ahead of print].

14. Segall G, Delbeke D, Stabin MG, et al. SNM practice guideline for sodium 18F-fluoride PET/CT bone scans 1.0. *J Nucl Med.* 2010;51:1813-1820.

15. Beheshti M, Mottaghy FM, Paycha F, et al. (18)F-NaF PET/CT: EANM procedure guidelines for bone imaging. *Eur J Nucl Med Mol Imaging.* 2015;42:1767-1777.

16. Cornford P, Bellmunt J, Bolla M, et al. EAU-ESTRO-SIOG Guidelines on prostate cancer. Part II: treatment of relapsing, metastatic, and castration-resistant prostate cancer. *Eur Urol.* 2017;71:630-642.

17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159-174.

18. Kurdziel KA, Shih JH, Apolo AB, et al. The kinetics and reproducibility of 18F-sodium fluoride for oncology using current PET camera technology. *J Nucl Med.* 2012;53:1175-1184.

19. Lin C, Bradshaw T, Perk T, et al. Repeatability of quantitative 18F-NaF PET: a multicenter study. *J Nucl Med.* 2016;57:1872-1879.

20. Rohren EM, Etchebehere EC, Araujo JC, et al. Determination of skeletal tumor burden on 18F-Fluoride PET/CT. *J Nucl Med.* 2015;56:1507-1512.

21. Lapa P, Marques M, Costa G, Iagaru A, Pedroso de Lima J. Assessment of skeletal tumour burden on 18F-NaF PET/CT using a new quantitative method. *Nucl Med Commun.* 2017;38:325-332.

22. Fendler WP, Calais J, Allen-Auerbach M, et al. (68)Ga-PSMA-11 PET/CT interobserver agreement for prostate cancer assessments: an international multicenter prospective study. *J Nucl Med.* 2017;58:1617-1623.

23. Fendler WP, Calais J, Eiber M, et al. Assessment of 68Ga-PSMA-11 PET accuracy in localizing recurrent prostate cancer: a prospective single-arm clinical trial. *JAMA Oncol.* 2019.

24. James ND, Sydes MR, Clarke NW, et al. Addition of docetaxel, zoledronic acid, or both to first-line long-term hormone therapy in prostate cancer (STAMPEDE): survival results from an adaptive, multiarm, multistage, platform randomised controlled trial. *Lancet.* 2016;387:1163-1177.

25. James ND, de Bono JS, Spears MR, et al. Abiraterone for Prostate Cancer Not Previously Treated with Hormone Therapy. *N Engl J Med.* 2017;377:338-351.

26. Parker CC, James ND, Brawley CD, et al. Radiotherapy to the primary tumour for newly diagnosed, metastatic prostate cancer (STAMPEDE): a randomised controlled phase 3 trial. *Lancet.* 2018;392:2353-2366.

27. Langsteger W, Balogova S, Huchet V, et al. Fluorocholine (18F) and sodium fluoride (18F) PET/CT in the detection of prostate cancer: prospective comparison of diagnostic performance determined by masked reading. *Q J Nucl Med Mol Imaging.* 2011;55:448-457.

28. Beheshti M, Vali R, Waldenberger P, et al. Detection of bone metastases in patients with prostate cancer by 18F fluorocholine and 18F fluoride PET-CT: a comparative study. *Eur J Nucl Med Mol Imaging.* 2008;35:1766-1774.

29. Wondergem M, van der Zant FM, van der Ploeg T, Knol RJ. A literature review of 18F-fluoride PET/CT and 18F-choline or 11C-choline PET/CT for detection of bone metastases in patients with prostate cancer. *Nucl Med Commun.* 2013;34:935-945.

30. Pyka T, Okamoto S, Dahlbender M, et al. Comparison of bone scintigraphy and (68)Ga-PSMA PET for skeletal staging in prostate cancer. *Eur J Nucl Med Mol Imaging.* 2016;43:2114-2121.

31. Zacho HD, Barsi T, Mortensen JC, et al. Prospective multicenter study of bone scintigraphy in consecutive patients with newly diagnosed prostate cancer. *Clin Nucl Med.* 2014;39:26-31.

32. Palmedo H, Marx C, Ebert A, et al. Whole-body SPECT/CT for bone scintigraphy: diagnostic value and effect on patient management in oncological patients. *Eur J Nucl Med Mol Imaging.* 2014;41:59-67.

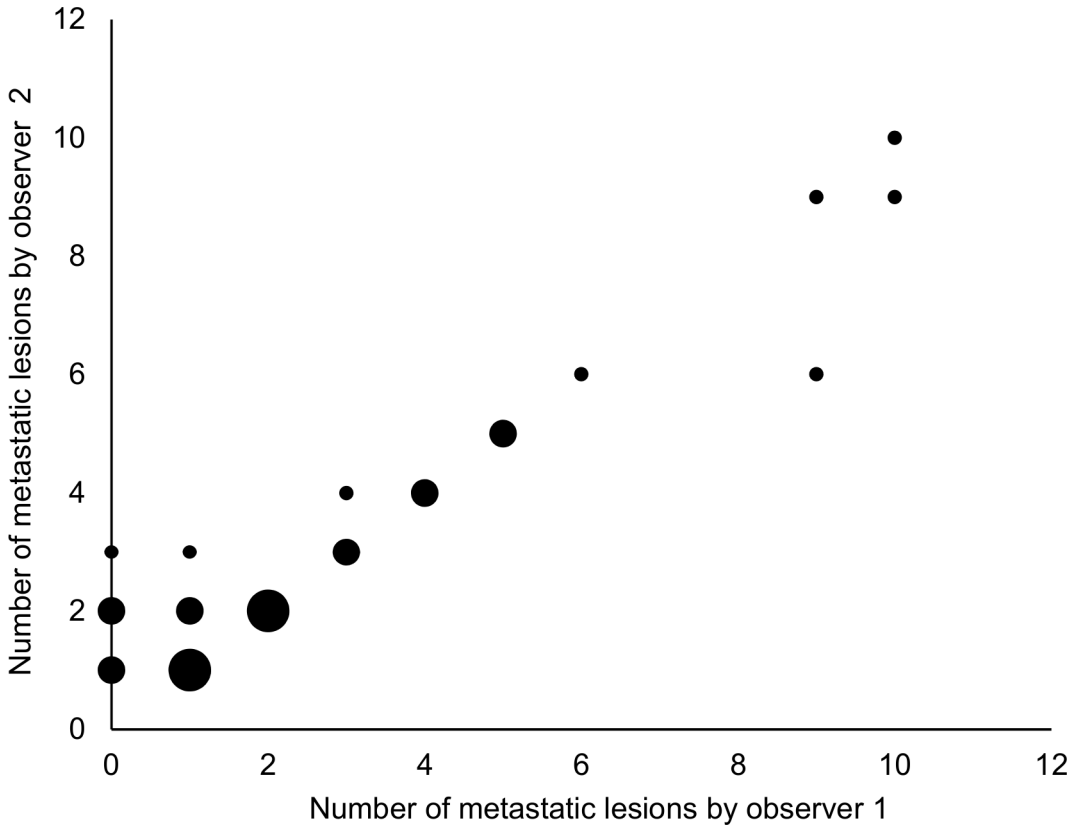
33. Zacho HD, Manresa JAB, Aleksyniene R, et al. Three-minute SPECT/CT is sufficient for the assessment of bone metastasis as add-on to planar bone scintigraphy: prospective head-to-head comparison to 11-min SPECT/CT. *EJNMMI Res.* 2017;7:1.

34. Hillner BE, Siegel BA, Hanna L, Duan F, Shields AF, Coleman RE. Impact of 18F-fluoride PET in patients with known prostate cancer: initial results from the National Oncologic PET Registry. *J Nucl Med.* 2014;55:574-581.

35. Gauthe M, Aveline C, Lecouvet F, et al. Impact of sodium (18)F-fluoride PET/CT, (18)F-fluorocholine PET/CT and whole-body diffusion-weighted MRI on the management of patients with prostate cancer suspicious for metastasis: a prospective multicentre study. *World J Urol.* 2018.

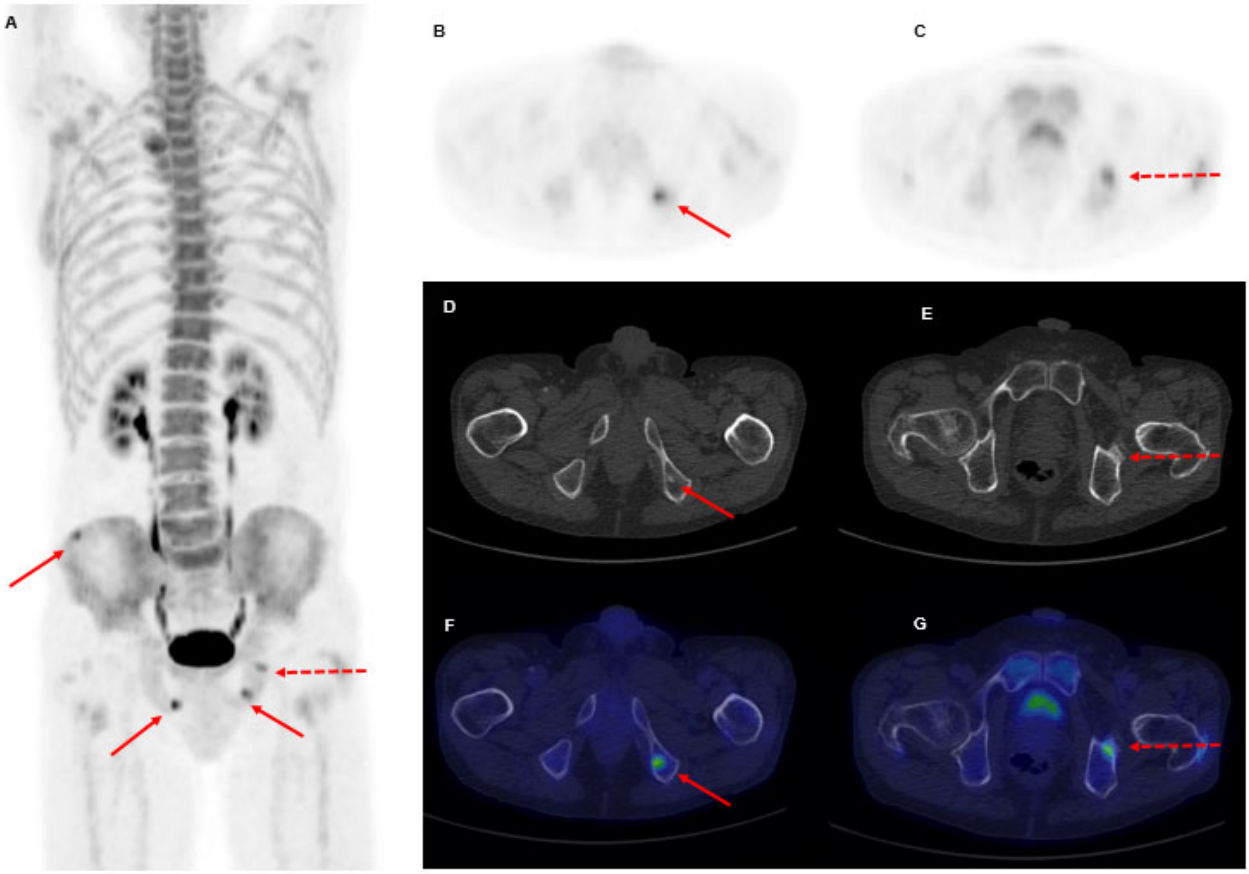
Figure legends

Figure 1.



Correlation between the number of bone metastases by each observer in the group of patients with 1 to 10 bone metastases as determined by at least one of the observers (n = 32). The size of the dots implies the number of identical values it represents. Small dots: 1 patient, medium dots: 2-3 patients, and large dots: 4-6 patients.

Figure 2



The maximum NaF PET intensity projection in a patient in whom both observers deemed the patient M1 (correctly according to the reference standard) but disagreed in the number of metastases (A). The observers agreed that the patient harbors three bone metastases marked with full arrows. In addition, one observer considered the lesion marked by the dotted arrow to be metastatic, whereas the other observer considered this a benign lesion. One lesion marked by the full arrow is shown in (B) as axial NaF PET-images, the corresponding sclerotic lesion in the CT image (D) and in the fused axial image (F). The lesion in which the observers disagreed is shown in (C) the axial NaF PET image

with the corresponding CT image (E), showing that the lesion is located close to the acetabular part of the hip. The fused image is shown in (G).

Table 1 Patient demographics and clinical characteristics

Patients, n	219
Age (years), mean (range)	68 (46-87)
PSA (ng/mL), median (range)	8.0 (0.2-9708)
Disease stage at the time of NaF PET/CT	
Staging	129
Biochemical recurrence	67
mCRPC	23
Gleason score	
< 7, n (%)	17 (8%)
7, n (%)	121 (55%)
8, n (%)	31 (14%)
9, n (%)	38 (17%)
10, n (%)	2 (1.0%)
Unknown, n (%)	10 (5%)
T-stage	
T1, n (%)	66 (30%)
T2, n (%)	86 (39%)
T3, *n* (%)	49 (22%)
T4, n (%)	11 (5%)
Unknown, n (%)	7 (4%)

PSA: Prostate-specific antigen; mCRPC: metastatic castration resistant prostate cancer.

Table 2

Agreement (95% confidence intervals) between two observers in different disease stages of prostate cancer.

		Stage of prostate cancer		
Three-category scale	**All patients**	**Staging, n = 129**	**BCR, n = 67**	**mCRPC, n = 23**
Unweighted Cohen's Kappa	0.83 (0.79-0.86)	0.82 (0.74-0.85)	0.70 (0.48-0.78)	1.00 (.-.)
Linear weighted Cohen's Kappa	0.90 (0.87-0.91)	0.89 (0.83-0.97)	0.79 (0.63-0.89)	1.00 (.-.)
Dichotomous scale				
p_{pos}	0.94 (0.90-0.98)	0.94 (0.87-0.99)	0.86 (0.67-1.00)	1.00 (.-.)
p_{neg}	0.97 (0.95-0.99)	0.97 (0.95-0.99)	0.97 (0.94-1.00)	1.00 (.-.)
Cohen's Kappa	0.91 (0.86-0.97)	0.91 (0.83-0.99)	0.83 (0.65-1.00)	1.00 (.-.)

BCR: Biochemical Recurrence, mCRPC: metastatic castration resistant prostate cancer.

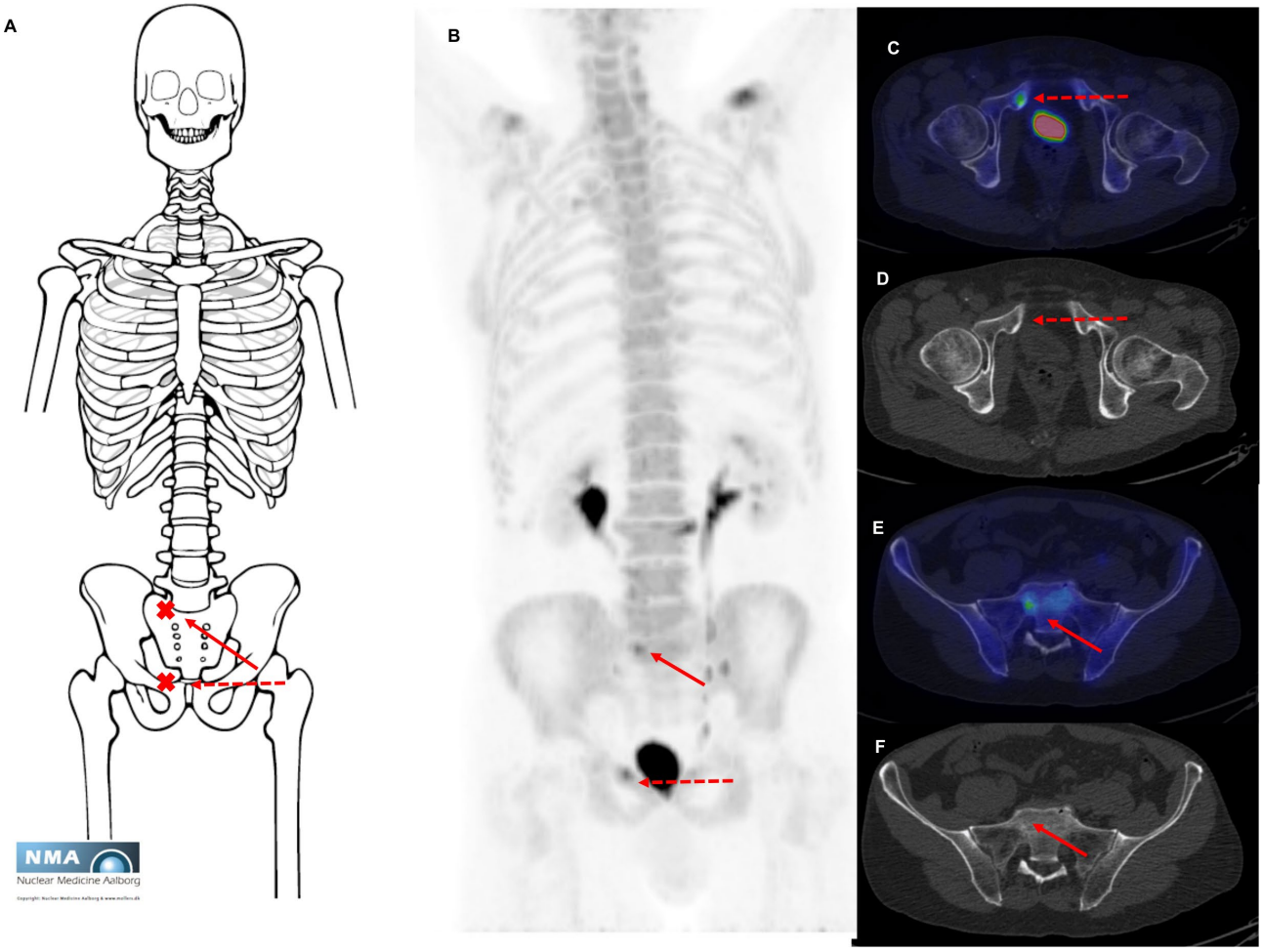
Table 3

Diagnostic accuracy of NaF PET/CT in 211 patients with a final diagnosis for the presence or absence of bone metastases. Data are shown in an optimistic analytical approach (pessimistic approach is written in italics).

	M0 by final diagnosis, n=147			M1 by final diagnosis, n=64						
Three point scale	**M0**	**Equivocal**	**M1**	**M0**	**Equivocal**	**M1**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**PPV (95% CI)**	**NPV (95% CI)**
Observer 1	136	7	4	5	4	55	0.86 (0.75-0.93) *0.92 (0.83-0.97)*	0.97 (0.93-0.99) *0.93 (0.87-0.96)*	0.93 (0.84-0.98) *0.84 (0.74-0.92)*	0.94 (0.89-0.97) *0.97 (0.92-0.99)*
Observer 2	122	21	4	5	2	57	0.89 (0.79-0.96) *0.92 (0.83-0.97)*	0.97 (0.93-0.99) *0.83 (0.76-0.89)*	0.93 (0.84-0.98) *0.70 (0.59-0.80)*	0.95 (0.91-0.98) *0.96 (0.91-0.99)*
Consensus by two observers	132	11	4	6	2	56	0.88 (0.77-0.94) *0.91 (0.81-0.96)*	0.97 (0.93-0.99) *0.90 (0.84-0.94)*	0.93 (0.84-0.98) *0.79 (0.68-0.88)*	0.95 (0.90-0.98) *0.96 (0.91-0.98)*
Dichotomous scale	**M0**		**M1**	**M0**		**M1**	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**PPV (95% CI)**	**NPV (95% CI)**
Observer 1	141		6	7		57	0.89 (0.79-0.96)	0.96 (0.91-0.98)	0.91 (0.80-0.96)	0.95 (0.91-0.98)
Observer 2	137		10	7		57	0.89 (0.79-0.96)	0.93 (0.88-0.97)	0.85 (0.74-0.93)	0.95 (0.91-0.98)
Consensus by two observers	142		5	8		56	0.88 (0.77-0.94)	0.97 (0.92-0.99)	0.92 (0.82-0.97)	0.94 (0.90-0.98)

M0: No bone metastases, M1: Bone metastasis present, PPV: positive predictive value, NPV: negative predictive value, CI: confidence interval. Consensus by two observers: In cases of disagreement, the observers re-evaluated the scans and reached a consensus diagnosis.

Supplemental Figure 1.



An example of disagreement on a patient level between the reviewers: One observer considered the patient M1 with two bone metastases, whereas, the other observer considered the same two lesions equivocal. The schematic drawing used to report the findings is shown in (A) with the two marks in red assigned by one of the observer and the corresponding maximum NaF PET intensity projection (B). The dotted arrow (in A and B) shows the lesion in the pubic bone with corresponding axial fused image in (C) and low dose CT in (D). The full arrow (in A and B) shows the lesion in the sacral bone with corresponding axial fused image (E) and ow dose CT (F). No corresponding changes in CT was seen for any of the lesions, which may in part explain the difference in interpretation. Illustration (A) is copyright of Nuclear Medicine Aalborg, printed with permission of Nuclear Medicine Aalborg.

Supplemental Table 1. Detailed description of the 19 subjects where the readers disagree on the skeletal status.

Patient number	Stage of PCa at the time of NaF PET/CT	PSA (ng/mL)	Age (years)	Gleason score	Clinical T-stage	NaF-PET/CT 3- point scale* Observer 1	NaF-PET/CT 3- point scale* Observer 2	Consensus of two observers 3-point scale*	Final diagnosis, patient level
1	Primary staging	7	62	7	cT1c	M0	Eqivocal, one lesion	Equivocal, one lesion	M0
2	Primary staging	11	71	7	cT1c	M0	Equivocal, one lesion	Equivocal, one lesion	M0
3	Primary staging	79	61	7	cT3b	Equivocal, one lesion	M1, one lesion	M1, one lesion	M1
4	BCR	0,3	75	7	cT1	Equivocal, one lesion	M1, two lesions	M1, one lesion	M1
5	Primary staging	5,8	62	7	cT1c	Equivocal, one lesion	M0	Equivocal, one leison	M0
6	Primary staging	11	73	7	cT2a	M0	Equivocal, one lesion	M0	M0
7	Primary staging	6,6	69	7	cT1c	M0	Equivocal, two lesions	M0	M0
8	Primary staging	6,6	69	7	cT1c	M0	Equivocal, one lesion	M0	M0
9	Primary staging	5,1	63	9	cT1c	M0	Equivocal, one lesion	M0	M0
10	Primary staging	6,1	61	7	cT1c	M0	Equivocal, two lesions	M0	M0

11	Primary staging	6,3	60	6	cT2b	M0	Equivocal, one lesion	M0	M0
12	BCR	0,9	60	8	cT2	M0	M1, one lesion	M1, one lesion	M0
13	BCR	5,7	78	7	cT2	M0	Equivocal, one lesion	M0	M0
14	BCR	2	72	7	cT2	M0	Equivocal, superscan or metabolic bone disease	M0	M0
15	BCR	0,8	72	7	cT1	M0	Equivocal, one lesion	M0	M0
16	Primary staging	17	68	7	cT1c	M1, two lesions	Equivocal, two lesions	Equivocal, two lesions	M0
17	BCR	0,6	55	6	cT1	M0	Equivocal, two lesions	M0	M0
18	BCR	3,3	71	7	cT3	M0	Equivocal, one lesion	Equivocal, one lesion	M0
19	Primary staging	10	72	7	cT1c	M0	Equivocal, one lesion	M0	M0

* Three-point scale: No bone metastasis, equivocal, bone metastasis present.
BCR: biochemical recurrence, M0: No bone metastases, M1: bone metastases present.