Journal of Nuclear Medicine, published on August 27, 2015 as doi:10.2967/jnumed.115.160085

Analytical Validation of the Automated Bone Scan Index as an Imaging Biomarker to Standardize the Quantitative Changes in Bone Scans of Patients with Metastatic Prostate Cancer

Aseem Anand[1,2], Michael J. Morris[1], Reza Kaboteh[3], Lena Båth[3], May Sadik[3], Peter Gjertsson[3], Milan Lomsky[3], Lars Edenbrandt[3,4], David Minarik[5], Anders Bjartell[2,6]

[1]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, United States .[2]Department of Translational Medicine Division of Urological Cancer, Lund University, Malmö, Sweden. [3]Department of Clinical Physiology, Sahlgrenska University Hospital, Gothenburg, Sweden.[4]Department of Clinical Physiology, Skåne University Hospital, Lund University, Malmö, Sweden. [5]Department of Radiation Physics, Skåne University Hospital, Lund University, Malmö, Sweden. [6]Department of Urological, Skåne University Hospital, Lund University, Malmö, Sweden.

Corresponding Author:

Aseem Anand

Department of Translational Medicine Division of Urological Cancer, Lund University

Waldenströms gata 5, SE 205 02 Malmö, Sweden

E-mail: aseem.anand@med.lu.se

Phone: +46706604084; Fax: +4640336911

Word count: 4601

Running Title – Analytical Validation of Bone Scan Index

Abstract

Rationale: A reproducible and quantitative imaging biomarker to standardize the evaluation of changes in bone scans is an unmet need for prostate cancer patients with skeletal metastasis. Here we have performed a series of analytical validation studies to evaluate the performance of the automated Bone Scan Index (BSI) as an imaging biomarker in patients with metastatic prostate cancer (mPCa).

Method: Three separate analytical studies were performed to evaluate accuracy, precision, and reproducibility of automated BSI. *Simulation Study:* Bone scan simulations with pre-defined tumor burdens were created to assess accuracy and precision. Fifty bone scans were simulated with a tumor burden ranging from low to high disease confluence (0.10 to 13.0 BSI). A second group of 50 scans was divided into 5 subgroups, each containing 10 simulated bone scans, corresponding to BSI values of 0.5, 1.0, 3.0, 5.0 and 10.0. *Repeat Bone Scan Study:* To assess the reproducibility in routine clinical setting, two repeat bone scans were obtained from mPCa patients after a single 600 MBq ^{99m}Tc MDP injection. *Follow-up Bone Scan Study:* Two follow-up bone scans of mPCa patients were analyzed to compare the inter-observer variability of the automated BSI with that of the qualitative visual reads in assessing changes between the bone scans. The automated BSI was calculated using the software EXINI bone^{BSI}. The results were evaluated using linear regression, Pearson's correlation, Cohen's kappa (κ) measurement, coefficient of variation and standard deviation (SD).

Result: Linearity of the automated BSI in the range of 0.10 to 13.0 was confirmed, and Pearson's correlation was observed at 0.995 (N=50, 95% CI 0.99–0.99, p<0.0001). The mean coefficient of variation was less than 20%. The mean BSI difference between the two repeat bone scans of 35 patients was 0.05 (SD=0.15), with an upper confidence limits at 0.30. The inter-observer agreement in the automated BSI was more consistent (κ=0.96, p <0.0001) than

the qualitative visual assessment of the changes (κ=0.70, p<0.0001) in bone scans of 173 patients.

Conclusion: The automated BSI is a consistent imaging biomarker with performance characteristics that can standardize the quantitative changes in bone scans of patients with mPCa.

Key words – Bone Scan Index (BSI), imaging biomarker, bone scan, metastatic prostate cancer, analytical validation

Background

Prostate cancer is a bone tropic cancer, and nearly 85% of patients with fatal prostate cancer are reported to have bone metastases (*1*). In clinical practice, a bone scan is the most prevalent and cost-effective diagnostic imaging tool to detect the onset of skeletal metastasis in advanced prostate cancer patients (*2*). However, the clinical utility of on-treatment changes in the bone scans of advanced metastatic prostate patients remains limited. This limitation is largely due to the lack of a consistent methodology to quantify changes in bone scans.

The Bone Scan Index (BSI), developed at the Memorial Sloan Kettering Cancer Center, is a fully quantitative analysis of skeletal metastasis in bone scans (*3*). The BSI represents the tumor burden in a bone scan as a percentage of the total skeletal mass and has shown clinical significance as a prognostic imaging biomarker (*4,5*). The labor-intensive process of manually calculating the BSI has prevented its wide-spread adoption in routine clinical practice.

The image analysis program developed by EXINI Diagnostics (Lund, Sweden), for the bone scan imaging modality, has automated the BSI calculation (*6-8*). The scan normalization and the iterative artificial neural network detect the abnormal hotspots that are suspected as metastatic lesions and generate BSI in a significantly shorter time-span (<10 seconds). In a subsequent

study, the automated BSI was shown to correlate with the manual BSI of newly diagnosed prostate cancer patients and it was independently associated with overall survival (*9*). However, the clinical qualification of the automated BSI as an imaging biomarker indicative of the treatment response hinges on its analytical performance characteristics, which are yet to be validated. The analytical validation of the automated BSI against a known analytical standard and against the pre-analytical variability of the routine bone scan procedure is essential to assess the automated BSI as a standardized quantitative platform for prospective clinical studies. Here, we have performed analytical studies that incorporate computer simulations and clinical patients to evaluate the accuracy, precision and reproducibility of the automated BSI. We hypothesized that with minimal manual supervision, the automated BSI can standardize the quantitative changes in bone scans of patients with metastatic prostate cancer.

Methods

Analytical study design: To evaluate the performance characteristics of the automated BSI, three separate analytical studies were performed. The pre-defined objectives and endpoint analysis for each of the three analytical studies are summarized in Table 1. The overall aim of these analytical studies was to test the hypothesis that automated BSI can standardize the quantitative changes in the bone scans. Ethical permission and individual patient consent were obtained to perform the analytical studies detailed below.

Simulation Study: The objective of the simulation study was to assess the accuracy and precision of the automated BSI against the known tumor burdens of the simulated bone scans. In the simulation study, two sets of fifty bone scans were simulated with known tumor burdens and their corresponding known phantom-BSI. The localization of the simulated tumor was randomized in all the bone scans. The first set of 50 simulated bone scans were created with focal tumor lesions ranging from low to high disease confluence (0.10 to 13.0 phantom-BSI).

Another set of 50 scans was divided into 5 subgroups, each containing 10 simulated bone scans, corresponding to phantom-BSI values of 0.5, 1.0, 3.0, 5.0 and 10.0.

Bone scan simulation: The SIMIND Monte Carlo (*10*) program together with the XCAT phantom (*11*), representing a standard male, were used to simulate bone scans with pre-defined tumor burdens in the skeleton. Randomly distributed focal lesions were inserted in the XCAT phantom skeleton by a MATLAB® script corresponding to a predefined BSI: phantom-BSI. The tumors were confined to the skeleton volume. The phantom-BSI was calculated as described in the original study (*3*). A restriction was set so that no tumor was placed below the mid-femur or below the mid-humerus. A virtual scintillation camera with a 9-thick-mm crystal, 9.5% energy resolution at 140 keV, a 256 x 1024 image matrix with a 2.4 mm pixel size, and a low-energy, high-resolution collimator was used for the simulations. To mimic real measurements, the simulations were performed with sufficient histories to avoid Monte Carlo noise. Instead Poisson noise was added after the simulations, corresponding to measurements with a total of 1.5 M counts in the anterior image. Anterior and posterior whole-body images were simulated for every phantom. The relative activity concentration in the bone, kidneys, bone marrow and tumors was set to 18, 9, 2.5 and 72, respectively, in relation to the remainder of the body.

Repeat Bone Scan Study: The objective of the repeat bone scan study was to determine the reproducibility threshold of the automated BSI in relation to the pre-analytical variability of the routine clinical bone scan procedure.

In this study, repeat whole-body bone scan was obtained from metastatic patients who were referred for a routine bone scan procedure. A previous bone scan was used to determine the presence of skeletal metastasis. The first whole-body bone scan was obtained after three hours of a single intravenous injection of 600 MBq technetium-99m methylene diphosphonate. The repeat bone scan was obtained directly after the completion of the first.

Patient bone scan: Whole-body images with anterior and posterior views (scan speed: 10 cm/min, 256 $\times$ 1024 matrix), were obtained using a gamma camera equipped with low-energy,

high-resolution parallel hole collimators (Maxxus; General Electric, Milwaukee, WI, USA). Energy discrimination was provided by a 15% window centered on the 140 keV of Tc-99m.

Follow-up Bone Scan Study: The objective of the follow-up bone scan study was to compare the inter-observer agreement of the visual reads with that of the automated BSI reads, in assessing changes between the two follow-up bone scans.

In the follow-up bone scan study, all prostate cancer patients with skeletal metastasis who underwent at least two whole-body examinations during the period from January 2002 to December 2008 as part of their clinical routine follow-up were considered for inclusion. Patients with digitally stored images were included, and in cases with more than two scans, the last two follow-up scans were used for the study.

The two follow-up bone scans for each patient were independently analyzed by three experienced nuclear medicine bone scan readers at three different read-sessions. At least three months elapsed between each read-session. In the first read-session, each reader independently classified the patients according to the signs of progressive metastatic disease. Patients with a follow-up scan showing new lesions and/or lesions present in the first scan that were distinctly larger were classified as having signs of progression. In the second read-session, the three readers independently classified all the patients based on the presence of two or more new lesions. In the third read-session, the three readers used the EXINI bone^{BSI} software to report changes in the automated BSI values between the two follow-up bone scans. The changes in the automated BSI were defined as an increase or decrease in the automated BSI greater than or equal to the reproducibility threshold obtained from the repeat bone scan study.

Automated BSI analysis*:* An upgraded EXINIbone^{BSI} version 2, developed by EXINI Diagnostics, was used to analyze the bone scans and to generate the automated BSI. The methodology of the automated platform has been described in detail in previous study (*9*). In

summary, the different anatomical regions of the skeleton are segmented followed by detection and classification of the abnormal hotspots as metastatic lesions. The weight fraction of the skeleton for each metastatic hotspot is calculated and the BSI is calculated as the sum of all such fractions.

Statistical analysis: In the simulation study, the automated BSI values were obtained for all the simulated bone scans and compared against the known standard, phantom-BSI, to assess analytical accuracy and precision. A linear regression model and Pearson's correlation test were used to evaluate the accuracy between the automated BSI values and that of the known phantom-BSI of the first set of 50 simulated bone scans. To confirm the method linearity, we tested the assumptions regarding residuals normality (Shapiro-Wilken test) and its homoscedasticity. Once the linear model was confirmed, the linear regression parameters were estimated, including a correlation coefficient (r), at a 95% confidence interval. In the second set of 50 simulated bone scans, the coefficient of variation and standard deviation were used to determine precision of the automated BSI at five different levels of the Phantom-BSI.

In the repeat bone scan study, the automated BSI was measured from both of the repeated bone scans. The difference between the two automated BSI reads was calculated to assess the reproducibility threshold against the pre-analytical variability of the routine clinical bone scan. The mean and standard deviation of the difference between the automated BSI reads were used to calculate the upper confidence limit, or the 95^{th} percentile value. This value represented the automated BSI reproducibility threshold for assessing consistent measurements of the changes in bone scans.

In the follow-up bone scan study, the inter-observer agreement of the visual subjective read against automated BSI was measured by Cohen's kappa agreement (κ), which measures agreement beyond that expected by chance. Kappa agreement was evaluated pairwise among the three readers—A vs. B, B vs. C and A vs. C—for each of the read-sessions. The mean

kappa agreement of the three paired readers was then compared with one another at each read session.

As an analytical observational study, no prior assumptions were made for the automated BSI performance to render power calculations. Statistical significance for each statistical test was set at 0.05. All statistical analyses were performed using IBM SPSS for Windows, Version 22.

Results

Simulation Study: The automated BSI, the dependent variable, was obtained from both sets of simulated bone scans and measured against the known phantom-BSI, which was considered the independent variable. In the first set of 50 simulated bone scans, the Shapiro-Wilk test confirmed that the residuals of the dependent variable were normally distributed (p=0.850). Additionally, the mean residual value of 0.00 with a standard deviation of 0.25 confirmed homoscedasticity showing constant variation across all values of the independent variable. Given that the residuals exhibited normality and homoscedasticity, the model was considered linear. The scatter plot with a linear fit line and the associated parameters for the linear regression in the range from 0.10 to 13.0 BSI are presented in Figure 1 and in Table 2, respectively. Pearson's correlation was observed to be 0.995 (95% CI: 0.99–0.99, p<0.0001). Table 3, provides the coefficient of variation and standard deviation of the automated BSI values at each of the predefined tumor burdens with varying localization for the second set of 50 simulated bone scans. The coefficient of variation at each of the five pre-defined phantom-BSIs was less than 20%.

Repeat Bone Scan Study: Thirty-five patients were consented and enrolled in the repeat bone scan study. All bone scans were eligible for automated BSI analysis. The Bland-Altman plot of the differences between the automated BSI reads of the 35 repeat bone scans is illustrated in Figure 2. The mean BSI difference between the two repeat bone scans was 0.05 with a

standard deviation of 0.15. The reproducibility threshold of the automated BSI for the consistent measurement of the changes in the bone scans, defined as the 95th percentile of the data, was observed at 0.30.

Follow-up Bone Scan Study: Bone scans of 173 metastatic prostate cancer patients were eligible for the study. The two consecutive bone scans from all 173 patients were independently analyzed by three experienced nuclear medicine readers for all three read-sessions. The pair-wise κ agreement at each read-session is demonstrated in Table 4. In read-sessions one and two, the mean κ agreement was 0.70 and 0.66 ($p < 0.0001$), respectively. In read-session three, the mean change in the automated BSI reads between the two follow-up bone scans of 173 patients was 1.84 (median=0.42, IQR:-2.4 to 10.26). The change in the automated BSI was defined as the increase or decrease beyond the reproducibility threshold of 0.30 BSI. Compared to the first two read-sessions, the mean κ agreement of the three readers in using automated BSI was observed much higher at 0.96 ($p < 0.0001$).

Discussion

Currently, there is no quantitative imaging biomarker in patients with metastatic prostate cancer. The on-treatment changes in bone scan are inadequately assessed in an interpreter-dependent visual analysis. The inherent variability of such assessment would affect the clinical association of the intended biomarker with clinical endpoints. The FDA's biomarker qualification review program explicitly states that the clinical validation of a biomarker is empirically significant only if the marker is measured consistently and reproducibly (*12*).

However, during the course of this study, we found that unlike those in the field of blood-based biomarkers, the guidelines to analytically validate a biomarker in image diagnostics are not as well defined. One of the most serious limitations, in the effort to analytically validate an imaging biomarker, has been the difficulty in procuring the true analytical standard: histological

confirmation of abnormal hotspots as metastatic lesions. Previous studies have circumvented the issue of the analytical standard unavailability by using a reference reader for the bone scan analysis (*13*). Such measure of using the manual assessment as a true analytical benchmark could limit the analytical validation to the skills of the reference individual. To be independent of such limitations, the analytical standard has to represent the known burden of disease against which the true performance of the biomarker can be evaluated.

In the absence of a true analytical standard, we believe that the simulated bone scans from XCAT phantoms with pre-defined tumor burdens, confined to the volume of the skeleton, and the SIMIND Monte Carlo simulation of the gamma camera could be an excellent alternative. The simulated bone scans with randomized tumor locations but with known tumor burden can create real patient scenarios to evaluate the performance of the automated BSI against a true analytical standard, which is the known ground truth.

Our results of the simulation study demonstrated that irrespective of the varying tumor localization, the accuracy and precision of the automated BSI were maintained from low to high disease confluence of focal lesions, in the given BSI range from 0.10 to 13.0. This result marks an improvement in the performance of EXINI's current version of the automated BSI compared to that of its predecessor. The first generation of EXINI's automated BSI platform was reported to underestimate the BSI values in patients with high disease burden (*9*). The consistent linearity of the upgraded version of the automated BSI is vital to its intended clinical utility in advance metastatic prostate cancer patients. With consistent accuracy and precision, the changes in the automated BSI would be reproducible and its association with clinical endpoints would be reliable.

In our simulated study, we also noted that the automated BSI platform has a blind spot for lesions superimposed on the bladder. To avoid classification of tracer activity from the urinary bladder, the lesions simulated in the lower sacrum, coccyx and pubic regions were sporadically not classified as metastatic lesions. An example is illustrated in Figure 3. The resulting

automated BSI value for such simulations was slightly lower than the true phantom-BSI. Given the technical limitation of the platform, manual supervision should be considered when analyzing the bone scans of patients with hotspots in the lower sacrum, coccyx or pubic region.

Although a true analytical standard, the simulated bone scans are limited in its representation of the variables associated with the patient bone scans. The simulated bone scans represent one single individual, rendering the same activity concentration and varying only by the tumor burden and the tumor localization. Therefore, unlike the simulation study, the purpose of the repeat bone scan study in patients with skeletal metastasis was to evaluate the reproducibility of the automated BSI in clinical environment, which is pre-disposed to the variables associated with routine bone scan procedure. The variables such as image count corresponding to the scanning time and the inter-patient dependent attenuation factors can cause noise in BSI reads. Such noise can affect the reliability of the on-treatment BSI change as a quantitative bio-marker. In previous studies, the on-treatment BSI change has been reported as the percentage of BSI change, BSI doubling time and BSI difference (*5,14,15*), but none of the studies have accounted for the noise in BSI due to the inherent variability of the bone scan procedure.

Although limited in scale, our study is the first attempt to empirically address the noise in the automated BSI associated with the pre-analytical variability of the routine bone scan procedure. We propose that accounting for the reproducibility threshold of 0.30 will result in a consistent and reliable on-treatment BSI change that is clinically relevant. The preliminary data presented here warrants further validation of this threshold and its clinical implications.

In the follow-up bone scan study, the read-sessions that incorporated visual assessment of the changes in bone scans reported higher discrepancies among the three readers. The three readers, in our study, belonged to the same hospital and their interpreting style for bone scan assessment were probably more consistent than if readers had been from different centers. Prior studies have reported similar discordance in visual assessment of bone scans. In a Swedish study, 37 readers from 18 hospitals showed considerable disagreement in the bone

scan assessment. The kappa agreement among the readers ranged between 0.16 and 0.82, with a mean of 0.48 (*8*).

Compared to the visual assessment of the changes in the bone scans, the inter-observer agreement among all the three nuclear medicine readers increased significantly using the automated BSI platform. The manual supervisions of the automated BSI assessment in bone scans with abnormal hotspot located at the lower sacrum, coccyx or pubic region were the cause of some limited ambiguities among the readers.

Further, the automated BSI represented the changes in the bone scans as a continuous numerical variable. The gradient of this change could potentially have clinical utility in assessing both response and progression. A confirmed sequential increase in the automated BSI value can not only augment the Prostate Cancer Working Group two (PCWG2) recommendation of tracking two or more new lesions as an endpoint for radiographic progression but the confirmed sequential decline in the treatment follow-up BSI value could also be evaluated as a biomarker indicative of efficacy response.

Despite the advantages of a fully quantitative imaging biomarker, the automated BSI assessment does not absolve the inherent limitations of bone scan as a non-tumor specific imaging modality. The uptake of technetium 99m in bone scan reflects the increase in the osteoblastic activity in bone. As a result, in response to an effective treatment, the bone scan can show a temporary increase in activity. Future clinical investigations are warranted to develop BSI guidelines, similar to those incorporated in PCWG2 for bone scan progression, to determine the on-treatment BSI change that is clinically relevant.

Conclusion

In summary, we have demonstrated that with minimal manual supervision the automated BSI overcomes the limitations of qualitative visual assessment by providing an accurate, precise and

reproducible platform to standardize the quantitative changes in the bone scans for prostate cancer patients with skeletal metastasis. This study is the foundation for subsequent clinical investigation aimed at validating the clinical utility of the changes in the automated BSI as a consistent quantitative imaging biomarker indicative of treatment response.

Disclosure:

Outside the submitted work, Professor Lars Edenbrandt is a shareholder and part owner of EXINI Diagnostics AB, Professor Anders Bjartell and Aseem Anand are consultants to EXINI Diagnostics AB. There are no other disclosures.

References

1. Jacobs SC. Spread of prostatic cancer to bone. *Urology*, 1983; 21:337-44.
2. Heidenreich A, Bastian PJ, Bellmunt J, et al. EAU guidelines on prostate cancer. part 1: screening, diagnosis, and local treatment with curative intent-update 2013*. Eur Urol.* 2014; 65:124-37.
3. Imbriaco M, Larson SM, Yeung HW, et al. A new parameter for measuring metastatic bone involvement by prostate cancer: the Bone Scan Index. *Clin Cancer Res*. 1998; 4:1765-72.
4. Sabbatini P, Larson SM, Kremer A, et al., Prognostic significance of extent of disease in bone in patients with androgen-independent prostate cancer*. J Clin Oncol*. 1999. 17:948-57.
5. Dennis ER, Jia X, Mezheristskiy IS, et al. Bone scan index: a quantitative treatment response biomarker for castration-resistant metastatic prostate cancer. *J Clin Oncol.* 2012; 30:519-24.
6. Sadik M, Suurkula M, Hoglund P, Jarund A, Edenbrandt L. Improved classifications of planar whole-body bone scans using a computer-assisted diagnosis system: a multicenter, multiple-reader, multiple-case study. *J Nucl Med*. 2009; 50:368-75.
7. Sadik M, Hamadeh I, Nordblom P, et al. Computer-assisted interpretation of planar whole-body bone scans. *J Nucl Med*. 2008; 49:1958-65.
8. Sadik M, Suurkula M, Hoglund P, Jarund A, Edenbrandt L. Quality of planar whole-body bone scan interpretations--a nationwide survey. *Eur J Nucl Med Mol Imaging*. 2008; 35:1464-72.
9. Ulmert D, Kaboteh R, Fox JJ, et al. A novel automated platform for quantifying the extent of skeletal tumour involvement in prostate cancer patients using the Bone Scan Index. *Eur Urol.* 2012; 62:78-84.

10. Ljungberg M, Strand SE. A Monte Carlo program for the simulation of scintillation camera characteristics. *Comput Methods Programs Biomed*. 1989; 29:257-272.

11. Segars WP, Sturgeon G, Mendonca S, Grimes J, Tsui BM. 4D XCAT phantom for multimodality imaging research. *Med Phys*. 2010; 37:4902-15.

12. Goodsaid F, Frueh F. Biomarker qualification pilot process at the US Food and Drug Administration. *AAPS J*. 2007; 9:E105-8.

13. Brown MS, Chu GH, Kim HJ, et al. Computer-aided quantitative bone scan assessment of prostate cancer treatment response. *Nucl Med Commun.* 2012; 33:384-94.

14. Armstrong AJ, Kaboteh R, Carducci MA, et al. Assessment of the bone scan index in a randomized placebo-controlled trial of tasquinimod in men with metastatic castration-resistant prostate cancer (mCRPC). *Urol Oncol*. 2014; 32:1308-16.

15. Kaboteh R, Gjertsson P, Leek H, et al. Progression of bone metastases in patients with prostate cancer - automated detection of new lesions and calculation of bone scan index. *EJNMMI Res*. 2013; 3:64-70.

Figures:

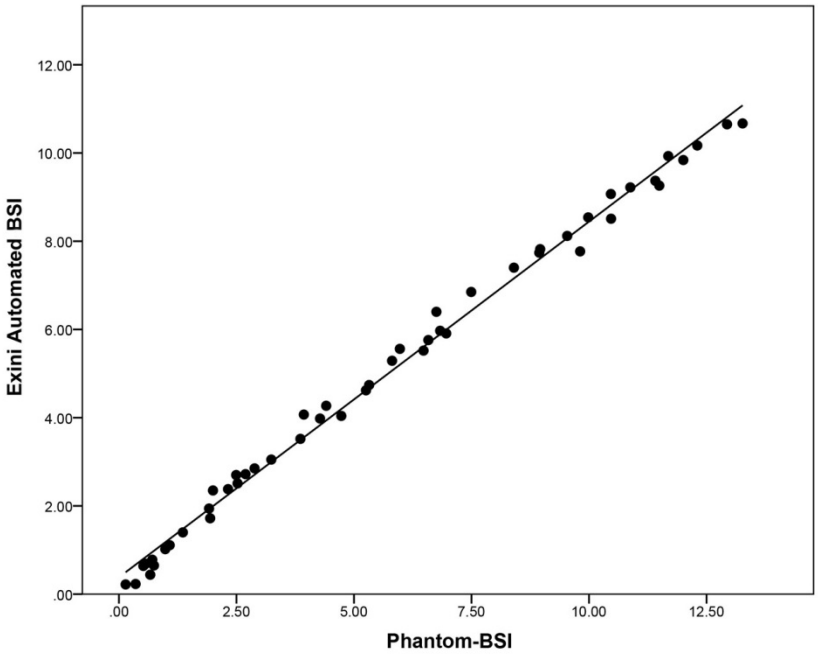

Figure 1. Linearity of the automated BSI against the first set of 50 simulated bone scans – Scatter plot of EXINI's automated BSI values against the phantom-BSI, the known analytical standard.

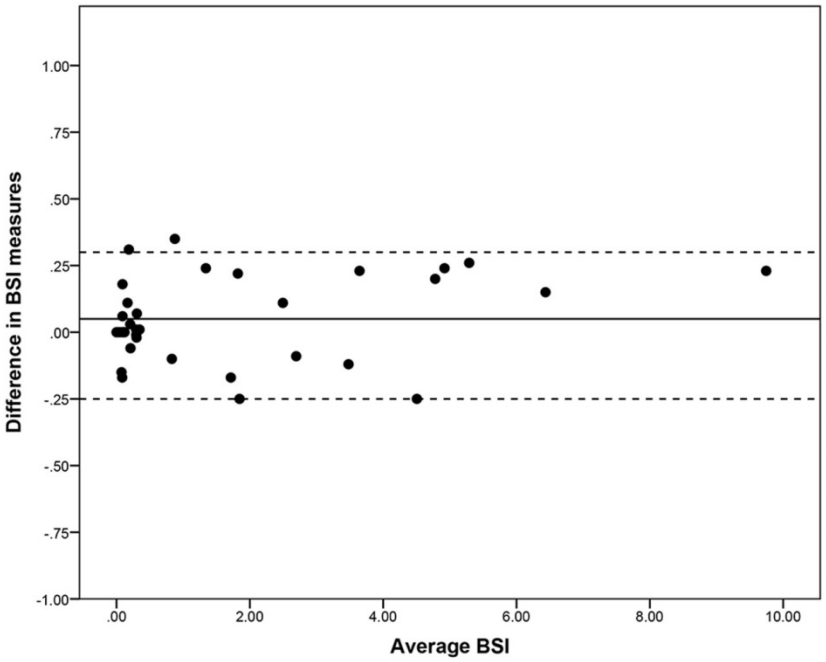

Figure 2. Bland-Altman plot to evaluate the reproducibility of the automated BSI reads from repeat bone scans of 35 metastatic patients. The mean BSI difference was 0.05 (solid horizontal line), and the standard deviation was 0.15, with an upper confidence limit of 0.30 and a lower confidence limit of -0.25 (horizontal dotted lines).

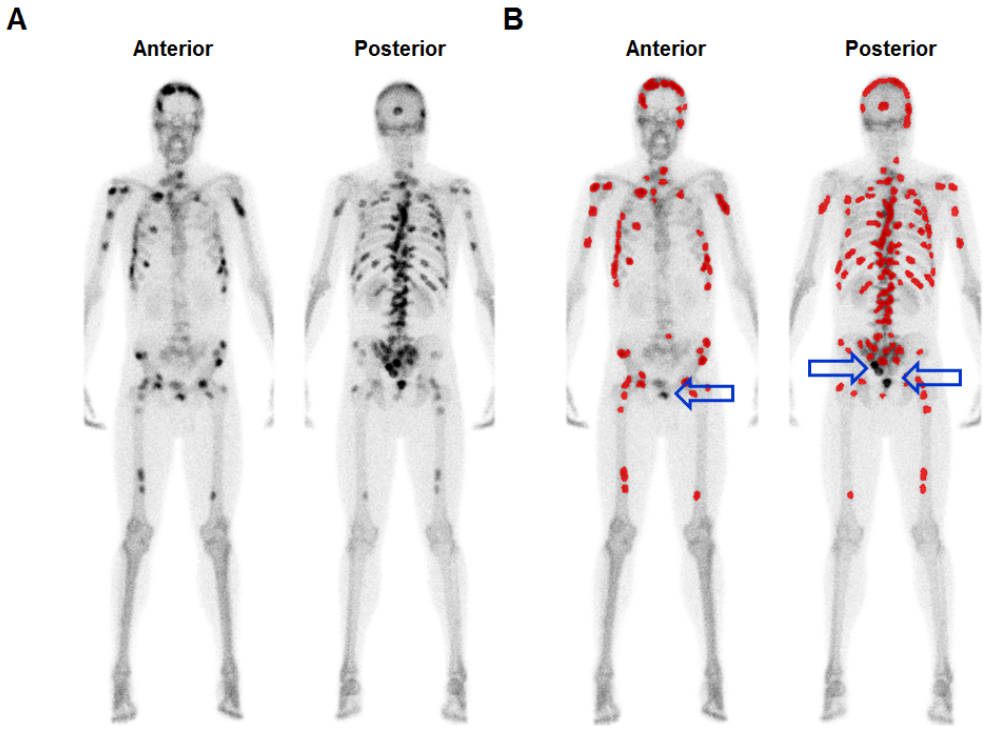

Figure 3. A. Example of a simulated bone scan (anterior and posterior image) with a known tumor burden corresponding to phantom-BSI 10.0. **B.** The lesions detected and classified as metastatic by the automated platform, for the BSI calculation, are highlighted in RED. The blue arrow indicates the simulated lesions in the lower sacrum that were not detected by the automated platform.

Tables:

Table 1. Summary of the analytical studies to evaluate the performance characteristics of the automated BSI as a consistent imaging biomarker that could standardize the quantitative analysis of bone scans in multi-institutional clinical studies.

Analytical Studies	Objectives	Design	Endpoint
1. Simulation Study	Accuracy and Precision	Simulation of bone scans with known Phantom-BSI as analytical standard	Measuring the automated BSI against Phantom-BSI
2. Repeat Bone Scan Study	Reproducibility	Metastatic patients with repeat bone scans	Measuring the difference of the two automated BSI reads
3. Follow-up Bone Scan Study	Inter-Observer variability	Metastatic patients with two routine follow-up clinical bone scans	Measuring the observer agreement in assessing automated BSI change

Table 2. The parameters for the linear regression model in the first set of 50 phantoms with the pre-defined BSI range of 0.10 to 13.0.

Linearity measures	Value	95% CI	Sig.
R	0.99	(0.99 – 0.99)	<0.0001
Slope	0.80	(0.78 – 0.83)	<0.0001
Intercept	0.38	(0.25 – 0.51)	<0.0001

Table 3. Standard deviation (SD) and coefficient of variation (CV) for the automated BSI in the second set of 50 phantoms at five pre-defined tumor burdens.

Phantom-BSI; N=50	**0.5 N=10**	**1.0 N=10**	**3.0 N=10**	**5.0 N=10**	**10.0 N=10**
SD	0.10	0.13	0.26	0.17	0.29
CV	0.19	0.11	0.08	0.06	0.03

Table 4. Pairwise Cohen's kappa agreement evaluating inter-observer agreement among the three independent readers to assess changes in the bone scans from 173 patients with metastatic prostate cancer.

Read-Sessions N=173	Pairwise Kappa Agreement		
	Reader A vs. B	Reader A vs. C	Reader B vs. C
1. Increased burden	0.56	0.90	0.65
2. Two new lesions	0.62	0.81	0.55
3. Change in BSI	0.96	0.97	0.96