Table of Contents

Address: 1111 Deming Way, Suite 105, Madison, WI 53717 **Phone**: 608-268-9684 **website**: www.aiq-solutions.com

AIQ Technology Validation Methodology

Document Version 1.0

Abstract

Introduction

A reproducible and quantitative imaging technique is needed to standardize the evaluation of changes in PET/CT scans of prostate cancer patients with skeletal metastasis. We performed a series of analytic validation studies to evaluate the performance of the semi-automated Bone Mets Application for patients diagnosed with metastatic prostate cancer.

Summary

Three analytic studies were performed to evaluate the accuracy, precision, and reproducibility of the Bone Mets application.

Performance Indicator	Studies	Complete	Verified
Accuracy	Simulation Study (Phantoms)	☒	☒
Reproducibility	Reproducibility Study 1	☒	☒
	Reproducibility Study 2	☒	☒
Precision	Reference Study 1	☒	☒
	Reference Study 2	☒	☒
Clinical Endpoints	Reference Study 3	☒	☒

Methods

Three analytic studies were performed to evaluate the accuracy, precision and reproducibility of AIQ Software Platform Bone Mets Application (BMA).

- **Accuracy Simulation Study:** PET/CT scan simulations, with predefined SUV uptake, were created to assess accuracy and precision. Two pairs of PET/CT scans were simulated with predefined tumors:

Address: 1111 Deming Way, Suite 105, Madison, WI 53717 **Phone**: 608-268-9684 **website**: www.aiq-solutions.com

- Phantom 1 consists of two scans, Scan 1 and Scan 2, and was designed to test detection thresholds and mathematical evaluation of change using static, predictable standardized Uptake Values (SUV). Regions of Interest (ROI) were placed in the scans and assigned static SUV values of 10 in the first scan, and 20 in the second scan.
- Phantom 2 consists of two scans, Scan 1 and Scan 2, and was designed to test variable thresholds and mathematical evaluation of change using a variety of SUV values in different parts of the anatomy. ROIs were defined ranging from low (1) to high (20) SUV and placed in a variety of bones.

- **Reproducibility Study 1:** PET/CT scan sets for four metastatic prostate cancer patients were analyzed multiple times, without manual intervention, to prove that the tool produces the same results and can handle a variety of skeletal variations.
- **Reproducibility Study 2:** PET/CT scan sets for four metastatic prostate cancer patients were analyzed manually by a qualified medical physics researcher and results were compared with the results of automated processing of the same scans, to prove that the results of the automated analysis are fundamentally the same as manual results produced by an imaging expert.

Results

Accurate localization, identification, segmentation and quantification of ROIs in the range of 1-20 SUV was confirmed. Running the tool multiple times on the same scan series results in the exact same analytical results. The automated tool produced very similar outcomes to the manual processing by a medical physicist.

Conclusion

The automated ROI localization, identification, segmentation and quantification provides a consistent imaging biomarker capable of standardizing quantitative changes in PET/CT scans of patients with metastatic prostate cancer. These references support limits of agreement used for ROI classification, and overall utility of BMA processes and tools.

Address: 1111 Deming Way, Suite 105, Madison, WI 53717 **Phone**: 608-268-9684 **website**: www.aiq-solutions.com

Overview

Prostate cancer is a bone-tropic cancer, and nearly 85% of patients with fatal prostate cancer are reported to have bone metastases [1]. PET/CT scans with NaF radiotracer are effective at finding and identifying ROIs, but the clinical utility of PET/CT scans is limited largely because of the lack of a reliable methodology to quantify ROIs and changes to ROIs in those scans.

The AIQ Software Platform Bone Mets Application (BMA), developed at the University of Wisconsin, is a semi-quantitative analysis of regions of tracer uptake in PET/CT scans. BMA produces statistical measures using Standardized Uptake Value to derive total uptake, maximum uptake, average uptake, heterogeneity of uptake, and PET active volume for the whole patient as well as for each ROI. SUV measurement using NaF PET/CT has shown clinical significance as a prognostic imaging biomarker [2-4]. However, the labor-intensive process of manually measuring and calculating statistics on regions of PET uptake has prevented widespread adoption in clinical practice.

The image-analysis software tool developed by AIQ Solutions, Inc. for BMA analysis automates localization, identification, segmentation and quantitative measurement using SUVs for ROIs. The software uses anatomy-specific SUV thresholding to detect hotspots, segments ROIs that are in bone, and calculates metrics. Previous BMA results have been independently associated with progression-free survival [3].

However, the clinical qualification of automated BMA as an imaging biomarker indicative of change depends on its analytic performance characteristics, which are yet to be validated. The analytic validation of automated BMA against a known analytic standard and against the variability of PET/CT procedures is essential to assess the automated BMA as a standardized quantitative platform for prospective clinical studies. We have performed analytic studies that incorporate simulations and clinical patients to evaluate the accuracy, precision and reproducibility of the semi-automated BMA. Our hypothesis is that with minimal manual supervision, semi-automated BMA could standardize the quantitative changes in bone scans of patients diagnosed with metastatic prostate cancer.

Address: 1111 Deming Way, Suite 105, Madison, WI 53717 **Phone**: 608-268-9684 **website**: www.aiq-solutions.com

Materials and Methods

Table 1

Summary of Analytic Studies to Evaluate Performance Characteristics of BMA as a Consistent Imaging Biomarker to Standardize Quantitative Analysis of PET/CT Scans

Analytic Study	Objective	Design	Endpoint
Simulation Study	Accuracy and Precision	Simulation of PET/CT scans with known phantom ROIs as analytic standard	Measuring automated BMA results against phantom ROIs
Reproducibility Study 1	Reproducibility of Automated Results	Metastatic patients with two scans spaced 13 weeks apart	Measuring difference between two automated BMA interpretations
Reproducibility Study 2	Reproducibility of Results	Metastatic patients with two scans spaced 13 weeks apart	Measuring difference between automated BMA and manual BMA analysis

Study Design

Three analytic studies were performed to evaluate the performance characteristics of automated BMA. The predefined objectives and endpoint analysis for each of the studies are summarized in **Table 1**. The overall aim of the studies was to test the hypothesis that BMA can standardize quantitative changes in PET/CT scans. Ethical permission and individual patient consent were obtained.

Simulation Study. The objective of the simulation study was to assess the accuracy and precision of BMA against the known tumor burdens of simulated PET/CT scans.

Address: 1111 Deming Way, Suite 105, Madison, WI 53717 **Phone**: 608-268-9684 **website**: www.aiq-solutions.com

In the simulation study, two phantoms were created with known tumor burdens and corresponding known BMA metrics. The locations of simulated tumors were established from existing patient scans but were manually manipulated to have specified SUV values throughout. A set of two male patient PET/CT scans were used as the starting point for each of the two phantom scan sets.

The first phantom scan set was designed to test thresholds and mathematical calculations by creating ROIs in various parts of the skeleton and setting the SUV value of each ROI to a static value in each scan in the series. All ROIs in the first timepoint scan were set to 20, and all ROIs in the second timepoint scan were set to 10. Based on regional thresholds, the software should categorize ROIs appropriately, and mathematical calculations and classification of ROI change was done manually based on the set values for reference comparison. These phantoms are shown in **Figure 1**.

The second phantom scan set was designed to test thresholds and mathematical calculations using a larger number of ROIs with more varied SUV values and more varied changes in SUV between the first and second scan. ROIs in the first and second scan sets were manually set to a variety of SUV values between 1 and 20 and distributed randomly within the skeleton. Mathematical calculations and classification of ROI change was done manually based on the set values for reference comparison. These phantoms are shown in **Figure 2**.

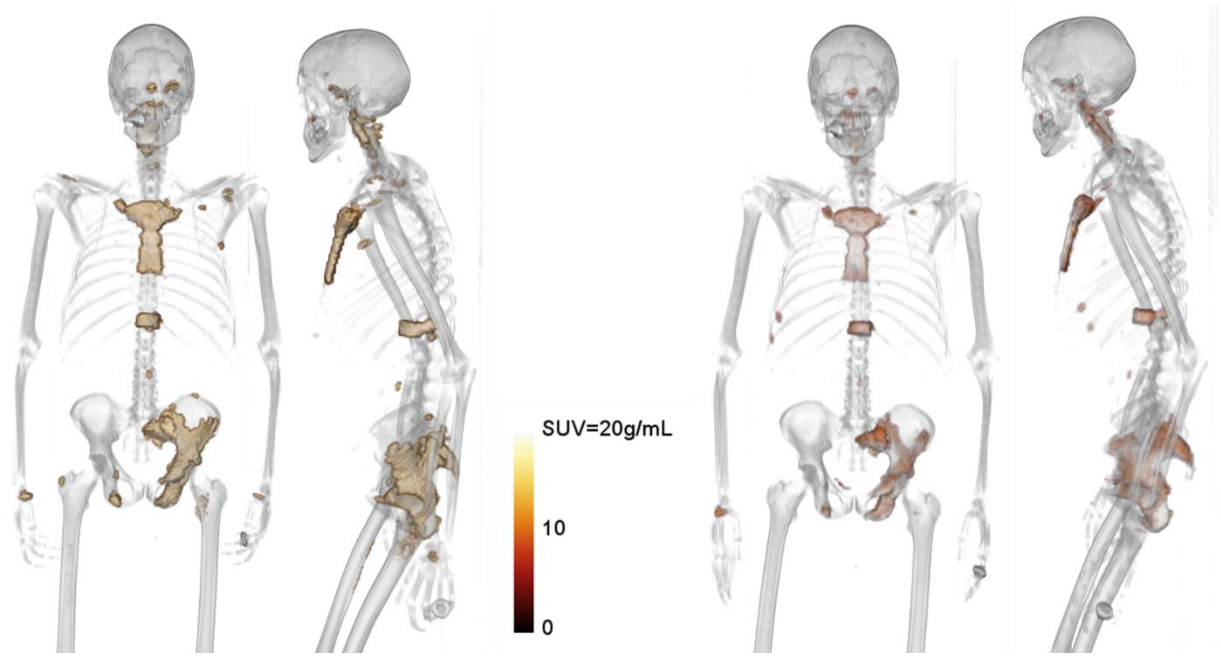


Figure 1: *Simulated phantom set with static uptake in ROIs*.

Address: 1111 Deming Way, Suite 105, Madison, WI 53717 **Phone**: 608-268-9684 **website**: www.aiq-solutions.com

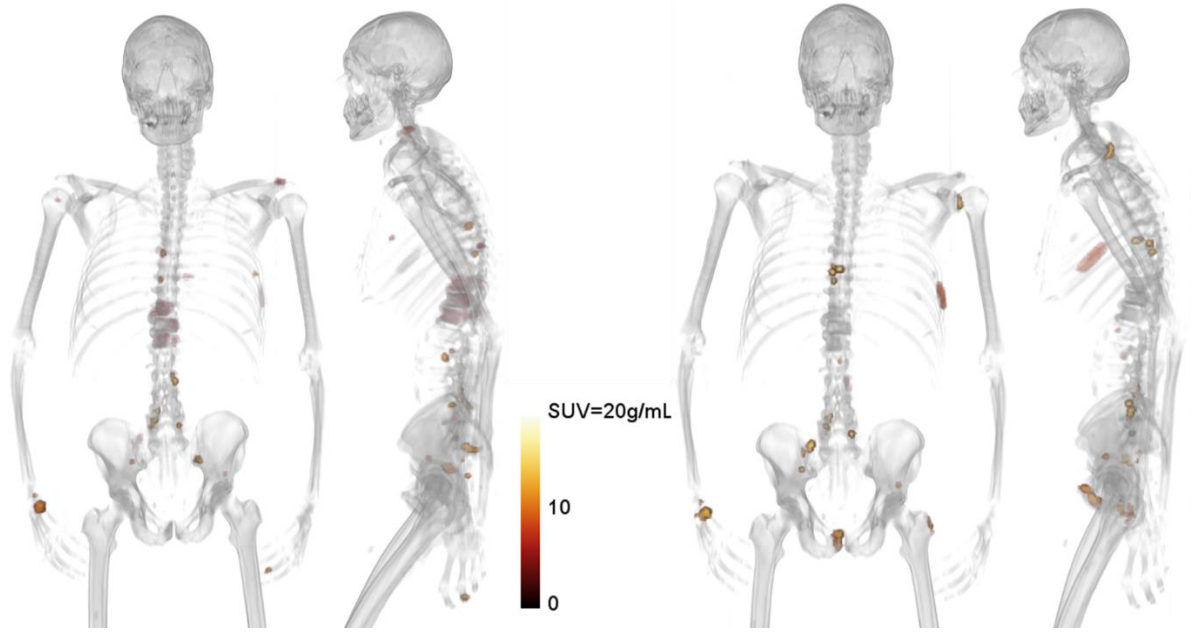
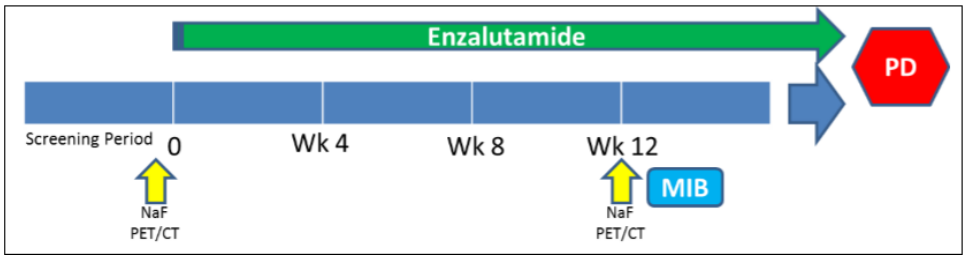


Figure 2: *Simulated phantom set with variable uptake in ROIs*.

Reproducibility Studies: patient population. Patients with metastatic prostate cancer with osseous metastases with plans to initiate treatment with enzalutamide were enrolled. All patients had identifiable bone lesions on 99mTc-MDP bone scintigraphy or NaF PET/CT. In addition, patients had to be able to comply with all study procedures, including having both the ability and willingness to lie flat for ≥ 30 minutes.

Image Capture. Before starting the therapy, all patients underwent NaF PET/CT scanning within 14 days prior to starting enzalutamide. All patients then underwent repeat NaF PET/CT scanning at week 12 (+/- 2 weeks) to assess change.



Reproducibility Study 1. The objective of the first reproducibility study was to establish that automated BMA analysis produces the exact same results on subsequent executions with no manual intervention. In this study,

Address: 1111 Deming Way, Suite 105, Madison, WI 53717 **Phone**: 608-268-9684 **website**: www.aiq-solutions.com

two analysists performed automated BMA analysis on each scan for the four patients. The BMA metrics of each scan were compared when obtained from the different analysts.

Reproducibility Study 2. The objective of the second reproducibility study was to establish that automated BMA analysis produces similar results to manual analysis by a trained medical physicist. The medical physicist operated using a set of scripts, rather than using an organized application and performed analysis on all of the patient images. BMA metrics obtained from each method were compared.

Semi-Automated Bone Mets Analysis

The semi-automated BMA module of QTxI version 1.0 was used to generate the BMA analysis results for all three studies. The methodology of the automated platform is described in detail in the QTxI User Manual. In summary, the PET scan is normalized to SUV, then ROIs are automatically located and identified based on the different anatomic regions of the skeleton and their bone density. They are then automatically segmented and quantified. Subsequent scans are registered to the first scan using skeletal segmentation and then ROIs are matched, and change is quantified. Finally, matched ROIs are classified based on statistical limits of agreement.

Results

Simulation Study

The BMA measurements (SUVmax, SUVmean, SUVtotal, Volume, and SUVstd(standard deviation) were obtained from each simulated PET/CT scan for individual ROIs and for the whole phantom, and change measurements (change in SUVmax, SUVmean, SUVtotal, Volume and SUVstd) were obtained from each set of simulated scans at a matched ROI and the whole phantom. These values were compared to the phantom BMA scan and change statistics as manually calculated during creation of the phantoms. All numbers matched precisely as expected.

Reproducibility Study 1: Automated Tool Reproducibility

BMA measurements (SUVmax, SUVmean, SUVtotal, Volume and SUVstd) were obtained from each patient PET/CT scan for individual ROIs and the whole patient. Change measurements (change in SUVmax, SUVmean, SUVtotal, Volume and SUVstd) were obtained from each set of patient scans at a matched ROI and whole patient.

Address: 1111 Deming Way, Suite 105, Madison, WI 53717 **Phone**: 608-268-9684 **website**: www.aiq-solutions.com

The measurements were obtained two times starting with import of the raw Digital Imaging and Communications in Medicine (DICOM) data and were run through the processing with no manual intervention for soft tissue removal or registration adjustment. All measurements matched precisely between the two runs.

Reproducibility Study 2: Process Reproducibility

BMA measurements (SUVmax, SUVmean, SUVtotal, Volume and SUVstd) were obtained from each patient PET/CT scan for individual ROIs and for the whole patient both by using the packaged and compiled QTxI software, and by a qualified medical physicist using the same process and scripts. Change measurements (change in SUVmax, SUVmean, SUVtotal, Volume and SUVstd) were obtained similarly. Measurements were compared between the semi-automated compiled software application and the manual execution of the process, shown in Table 2. All SUV metrics were found to be nearly identical between the different analyses with on average less than a 3% difference in values. Volume had a slightly larger average difference at 9%. These differences were largely due to the differences in the skeletal segmentations, which were found to be more accurate in the compiled QTxI software.

Table 2

Average percent difference in metrics obtained from semi-automated compiled QTxI or manual extraction by the medical physicist

Metric	SUV_{max}	SUV_{mean}	SUV_{total}	SUV_{hetero}	Volume
Average Percent Difference	0%	-1%	3%	-3%	9%

Clinical Endpoints Validation

Research versions of the AIQ Software Platform Bone Mets Application operated under the name Quantitative Total Bone Imaging (QTBI).

Address: 1111 Deming Way, Suite 105, Madison, WI 53717 **Phone**: 608-268-9684 **website**: www.aiq-solutions.com

Three clinical studies have been conducted using the same software scripts and process that are in the QTBI software tool. These studies provide data that supports the limits of agreement used for ROI classification in QTBI, link total ROI burden with long-term survival and as a result establishing the performance of the semi-automated detection process and tools, and establish statistical accuracy of the "Statistically Likely Benign" scoring methodology.

1] Lin C, Bradshaw T, Perk T, et al. Repeatability of quantitative 18F-NaF PET: a multicenter study. *J Nucl Med*.2016;57(12):1872-1879

- This publication assessed the repeatability of NaF PET imaging using QTBI. In this multicenter trial study, 35 patients with known bone metastases had test-retest whole-body NaF PET/CT scans repeated 3 ± 2 days apart, before the start of treatment. Patients were injected with a bolus of 111-185 MBq (3-5 mCi) of ^{18}F-NaF and imaged 60 minutes post-injection for 3 min per bed position from feet to skull vertex, using a standardized image-acquisition protocol. ROIs were identified and segmented using QTBI, followed by ROI verification by a nuclear medicine physician. Corresponding ROIs were automatically matched using QTBI's articulated registration from the two scans. QTBI metrics were extracted for each ROI and for the patient as a whole using the same methods described in our 510(k) version. Bland-Altman analysis was performed to assess the 95% limits of agreement to determine what changes in these metrics would be significant. The limits of agreement used in our 510(k) QTBI for hot spot classification (increasing or decreasing) is based on these findings.

2] T. Perk, T. Bradshaw, S. Chen, H. Im, S. Cho, S. Perlman, G. Liu and R. Jeraj 2018 Automated classification of benign and malignant lesions in ^{18}F-NaF PET/CT images using machine learning Phys. Med. Biol. 63 (2018) 225019

- 18F-NaF PET/CT imaging of bone metastases is confounded by tracer uptake in benign diseases, such as osteoarthritis. The goal of this work was to develop an automated bone lesion classification algorithm to classify lesions in NaF PET/CT images. Methods. A nuclear medicine physician manually identified and classified 1751 bone lesions in NaF PET/CT images from 37 subjects with metastatic castrate-resistant prostate cancer, 14 of which (598 lesions) were analyzed by three additional physicians. Lesions were classified on a five-point scale from definite benign to definite metastatic lesions. Classification agreement between physicians was assessed using Fleiss' **K**. To perform fully automated lesion classification, three different lesion detection methods based on thresholding were assessed: SUV > 10 g ml−1, SUV > 15 g ml−1, and a statistically

Address: 1111 Deming Way, Suite 105, Madison, WI 53717 **Phone**: 608-268-9684 **website**: www.aiq-solutions.com

optimized regional thresholding (SORT) algorithm. For each ROI in the image, 172 different imaging features were extracted, including PET, CT, and spatial probability features. These imaging features were used as inputs into different machine learning algorithms. The impact of different deterministic factors affecting classification performance was assessed. Results. The factors that most impacted classification performance were the machine learning algorithm and the lesion identification method. Random forests (RF) had the highest classification performance. For lesion segmentation, using SORT (AUC = 0.95 [95%CI = 0.94–0.95], sensitivity = 88% [86%–90%], and specificity = 0.89 [0.87–0.90]) resulted in superior classification performance ($p < 0.001$) compared to SUV > 10 g ml^{-1} (AUC = 0.87) and SUV > 15 g ml^{-1} (AUC = 0.86). While there was only moderate agreement between physicians in lesion classification (**K** = 0.53 [95% CI = 0.52–0.53]), classification performance was high using any of the four physicians as ground truth (AUC range: 0.91–0.93). Conclusion. We have developed the first whole-body automatic disease classification tool for NaF PET using RF, and demonstrated its ability to replicate different physicians' classification tendencies. This enables fully-automated analysis of whole-body NaF PET/CT images.

3] S. A. Harmon, T. Perk, C. Lin, J. Eickhoff, P. L. Choyke, W. L. Dahut, A. B. Apolo, J. L. Humm, S. M. Larson, M. J. Morris, G. Liu and R. Jeraj 2017 Quantitative Assessment of Early [18F]Sodium Fluoride Positron Emission Tomography/Computed Tomography Response to Treatment in Men With Metastatic Prostate Cancer to Bone J Clin Oncol 35 24 2829-2837 2017/06/28 Jun 27 Quantitative Assessment of Early [18F]Sodium Fluoride Positron Emission Tomography/Computed Tomography Response to Treatment in Men With Metastatic Prostate Cancer to Bone 1527-7755

- This publication used QTBI metrics extracted at multiple time points to assess NaF PET as a correlate to progression free survival (PFS). This study assessed 56 patients undergoing either standard taxane-based chemotherapy or androgen receptor pathway inhibitors, in which treatment benefit was being defined using standard imaging (e.g. bone scintigraphy, computed tomography scans) and clinical (e.g. symptoms, change in serum PSA levels). Patients enrolled on this study had ^{18}F-NaF PET/CT scans obtained at baseline and after 3 cycles of therapy (after 9-12 weeks) using the same image acquisition protocol as above. Hotspot detection was performed using QTBI, which was manually verified by a nuclear medicine physician. Using the same metrics provided by QTBI in this 510(k), global imaging metrics, including maximum standardized uptake value (SUV_{max}) and total functional burden (SUV_{total}), were extracted from composite ROI-level statistics for each patient. Progression-free survival (PFS) was calculated as a composite endpoint of progressive events using conventional imaging and/or physician discretion of clinical benefit; NaF imaging was not used for clinical evaluation. Cox

Address: 1111 Deming Way, Suite 105, Madison, WI 53717 **Phone**: 608-268-9684 **website**: www.aiq-solutions.com

proportional hazards regression analyses were conducted between imaging metrics and PFS. Using the limits of agreement obtained from the first supportive study, patients were stratified into groups, which was highly correlated to PFS.

Conclusion

We have demonstrated that with minimal manual supervision the automated BMA process overcomes the limitations of qualitative visual assessment and provides an accurate, precise and repeatable platform for standardizing quantitative changes in PET/CT scans of prostate cancer patients with detected ROIs. This study is the foundation for subsequent clinical investigations aimed at validating the clinical utility of changes in the BMA measurements as a consistent, quantitative imaging biomarker indicative of treatment change.

Address: 1111 Deming Way, Suite 105, Madison, WI 53717 **Phone**: 608-268-9684 **website**: www.aiq-solutions.com