

Comparative analysis of quality control results of 22 dual-energy X-ray absorptiometry (DEXA) systems across six manufacturers using a comprehensive one-day protocol

Fatemeh Abbasi¹, Reza Faghihi¹, Seyed Mohammad Entezarmahdi², Farshid Gheisari², Mahdi Haghighatafshar²

¹Department of Nuclear Engineering, Shiraz University, Shiraz, Iran

²Department of Nuclear Medicine, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran

Corresponding Author:

Dr. Seyed Mohammad Entezarmahdi

Address: Department of Nuclear Medicine, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran.

Email: mohammad.entezarmahdi@gmail.com

Running title: Comparative QC of 22 DEXA scanners: A one-day protocol

Article History:

Received: 10 December 2025

Revised: 31 August 2026

Accepted: 05 September 2026

Published Online: 28 September 2026

ABSTRACT

Introduction: International societies involved in bone mineral densitometry emphasize the necessity of regular quality control (QC) for dual-energy X-ray absorptiometry (DEXA) systems. However, the current recommended procedures are not comprehensive and do not assess certain critical performance aspects—such as system stability under varying soft-tissue thickness and differing attenuation coefficients—that may affect measurement reliability. This study introduces and applies a one-day QC protocol designed to evaluate five key aspects of DEXA performance: in-vitro short-term precision (IV-STP), repeatability, stability across varying soft-tissue thickness (St-Thi), stability across differing attenuation coefficients (St-Att), and accuracy.

Methods: Twenty-two DEXA scanners from six manufacturers were evaluated using a lumbar phantom with Plexiglass slabs or a water tank to simulate soft-tissue variations. IV-STP, repeatability, St-Thi, and St-Att were expressed as coefficients of variation (%CV), while accuracy - defined here as phantom-based agreement with the reference BMD value - was reported as the complement of the mean absolute error (%MAE).

Results: %CV values for IV-STP and repeatability ranged from 0.50–1.49 and 0.61–1.24, respectively. St-Thi and St-Att tests showed %CV ranges of 0.93–6.65 and 0.32–3.87. Phantom-based accuracy varied between 81.48% and 92.75%. These results indicated heterogeneity and suboptimal stability and accuracy among some systems.

Conclusion: The proposed five-step QC protocol revealed performance variations among DEXA systems, including aspects not addressed in existing recommendations by international societies.

This structured approach provides a more comprehensive framework for routine QC and can uncover latent deficiencies in system stability and accuracy, thereby enhancing diagnostic reliability and supporting regulatory standardization.

Keywords: Bone mineral densitometry; Dual-energy X-ray absorptiometry; Quality Control; Stability; Precision

INTRODUCTION

Osteoporosis is a major global public health issue [1, 2]. Dual-energy X-ray absorptiometry (DEXA) bone densitometry is the established “gold standard” for diagnosing and evaluating osteoporosis treatment [3]. DEXA provides high precision, short scan times, low radiation dose, and stable calibration, making it the World Health Organization’s (WHO) recommended method for assessing bone mineral density (BMD) [3, 4].

Before performing a DEXA scan, proper device calibration is essential. Calibration ensures accurate measurement by setting the correct voltage, filters, and detector functions, and by verifying that the software correctly identifies bone regions [5]. After calibration, quality control (QC) is required to confirm system performance. QC primarily evaluates precision, repeatability, and accuracy to ensure confidence in BMD results [6].

Precision refers to the system’s ability to reproduce consistent results under identical conditions [6]. The International Society for Clinical Densitometry (ISCD) recommends both *in vivo* and *in vitro* precision studies. *In vivo* precision is typically determined by repeating BMD tests three times on 15 patients or twice on 30 patients, calculating individual standard deviations, and deriving the group’s root mean square standard deviation. *In vitro* precision utilizes a standard QC phantom for measurements. Short-term precision is usually evaluated by scanning patients or phantoms on the same day, with repositioning between scans.

Repeatability assesses the consistency of BMD values in consecutive scans of a fixed, motionless object [6]. It reflects random variations in the tube, detector, and circuitry [6]. Since the scanned object remains unchanged, repeatability is expected to yield better results than precision tests.

Ideally, the thickness and composition of soft tissue should not affect the measured BMD value. However, discrepancies can occur due to beam hardening effects inherent in DEXA systems [7]. Different manufacturers apply various correction methods to minimize this influence, aiming to keep BMD results independent of soft tissue thickness and composition [7]. These corrections are calibrated using step phantoms, but their performance needs to be evaluated. Consequently, two *in vitro* stability tests have been introduced in the literature: stability across varying soft tissue thickness (St-Thi) [7] and stability across varying soft tissue composition (St-Att) [8].

The St-Thi test involves repeated measurements of a fixed bone phantom, with each scan conducted under different equivalent soft tissue thicknesses [7]. Ideally, BMD variations across these measurements should remain clinically and statistically insignificant.

The St-Att test evaluates stability when soft tissue composition changes while thickness remains constant. Using a fixed phantom, scans are repeated after altering the composition—typically the ratio of fat to lean mass—above the bone region [9]. This change affects the attenuation coefficient of soft tissue defined in the DEXA calibration. Since BMD is calculated using data from the bone region and its adjacent area, which normally includes only lean tissue of equal thickness, variations in soft tissue composition should not alter the BMD value [9].

Accuracy, defined by the International Standard Organization (ISO), is the degree to which a measurement agrees with an accepted reference value [6]. Assessing *in vivo* accuracy is difficult since it requires cadaver testing, which is rarely feasible [6]. Therefore, physical phantoms are widely used to assess the accuracy of DEXA systems. In the present study, accuracy was operationally defined as the percent agreement between phantom-measured BMD and the manufacturer-assigned reference BMD of the lumbar phantom, rather than a clinically validated accuracy against a gold-standard biological reference. The measured PA spine accuracy has previously been reported to be higher than 90% in the literature [10, 11].

This definition is consistent with the IAEA protocol for verifying DXA factory accuracy, in which the percent deviation between the phantom-measured average BMD and the manufacturer's reference BMD is used as the accuracy criterion [10]. Our %MAE-based accuracy metric is mathematically equivalent to 100% minus this percent deviation. We adopted the complement notation to express results as an intuitive percent-agreement figure, consistent with prior phantom-based DXA accuracy literature [10, 11], thereby enabling direct comparison across studies and manufacturers.

Although each manufacturer provides its own QC procedure, these protocols do not fully assess all aspects of DEXA system performance. The absence of a standardized and universally applicable QC method reduces the reliability of the reported results [12]. Thus, developing a comprehensive QC procedure suitable for all DEXA systems is essential. Such a procedure should be time-efficient, simple, and based on accessible, stable materials. Employing portable QC tools would also allow performance testing across multiple BMD centers.

In this study, we developed and implemented a one-day QC protocol to assess DEXA system performance. The protocol includes five stages: *in vitro* short-term precision (IV-STP), repeatability, St-Thi, St-Att, and accuracy. A lumbar phantom, combined with Plexiglass slabs or a water tank, was used to simulate various soft tissue thicknesses and attenuation coefficients. The proposed QC method was tested across 22 BMD centers equipped with DEXA systems from six manufacturers: Hologic, Norland, Lunar, Stratos, BM Tech, and DEXXUM-T.

METHODS

Phantom

In this study, a semi-anthropomorphic lumbar phantom, manufactured by Norland Company, was utilized. The phantom, a cuboid measuring 15 cm × 16 cm × 4 cm, is composed of a bone-like structure made from hydroxyapatite plus 603.79 grams of lean-like and 437.23 grams of fat-like tissue. The simulated bone mass density for the total L1-L4 area is 0.858 ± 0.01 gr/cm². This BMD value corresponds to a T-score of -1.5 within the Caucasian or Non-Hispanic white race database of Norland systems. Figure 1 presents a view of the Norland lumbar phantom.

To emulate varying soft tissue thickness above the lumbar, two types of tissue-equivalent materials were employed. The first consists of ten Plexiglas slabs, each with dimensions of 15 cm × 16 cm × 2 cm. The second is a plastic tank filled with water, measuring 15 cm × 16 cm × 16 cm. These tissue-equivalent phantoms were positioned above the lumbar phantom. Figure 2 illustrates the placement of the phantom on the DEXA table with the tissue-equivalent Plexiglas slabs and water tank. This Phantom was built at Shiraz University [13].

DEXA scanning

The DEXA measurements of the different phantom configuration were carried out with 22 different DEXA instruments. These systems are from six different manufacturer which are available in Iran market. The manufacturer companies and the DEXA system models is presented in Table 1. All instruments were calibrated and the recommended daily quality control test or calibration, as advised by the manufacturer, was performed on all devices at the beginning of the workday.

The scanning procedures for each of the DEXA instruments were similar. Each phantom array was positioned in the middle of the scan table. The scan modes used in this study conformed to the standard applications in each DEXA center. In those systems which have different analysis mode according with the tissue thickness, the scanning mode was set on the standard mode.

The software performed bone delineation, successfully detecting lumbar margins in most instances. In a few cases where automatic edge detection was poor, the operator manually adjusted the ROI. An expert nuclear medicine physician verified the accuracy of all ROIs and subsequent analyses. For QC procedures, the evaluated BMD values were derived from the L1-L4 region to minimize the impact of potential errors in bone region delineation.

Evaluation metrics and statistical analysis

In this study, the results of the IV-STP, repeatability, St-Thi, and St-Att tests are reported as percentages of the coefficient of variation (%CV). Concurrently, the results for phantom-based accuracy are reported as the complement to percentage mean absolute error (%MAE).

When consecutive measurements are performed on the phantom, the standard deviation (SD) of these consecutive scans can be calculated using Equation 1:

$$SD = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (1)$$

In this equation, n represents the number of consecutive measurements, x_i is the value of the i^{th} consecutive measurement, and $\bar{x}$ is the mean value of all n consecutive measurements.

The relative standard deviation (RSD) of the consecutive scans is then calculated using Equation (2):

$$RSD = \frac{SD}{\bar{x}} \quad (2)$$

Upon calculating the RSD, the %CV of a repeating test is then calculated using Equation (3):

$$(\%) CV = \sqrt{\frac{1}{m} \sum_{i=1}^m RSD_m^2} * 100 \quad (3)$$

Here, m is the number of repeating test sessions (each consisting of n consecutive measurements), and RSD_m is the relative standard deviation calculated for the m-th session using Equation 2. To compare the results of this study with those of a previously published paper, we have chosen to use the %CV rather than the least significant change (LSE) parameter.

The complement to percentage mean absolute error (%MAE) is calculated using Equation (4):

$$\text{Accuracy} = \text{Complement to MAE (\%)} = 100 - \frac{|BMD_1 - BMD_2|}{BMD_1} \times 100 \quad (4)$$

In the accuracy test, BMD_1 refers to the adjusted true BMD value reported on the phantom sheet, and BMD_2 refers to the mean of measured BMD values over the range of stable response area (further explained in the subsequent sections).

Normality of the data was assessed using Shapiro-Wilk test. Mean $\pm$ standard deviation (SD) of BMD were calculated. One sample t-tests were utilized to test the accuracy test. Shewhart charts were created to analyze the measured BMD over the soft tissue thickness change.

For precision assessment (%CV), we referenced the ISCD 2023 Position Statement threshold of less than 1.9% for in-vivo spinal BMD precision as a benchmark for contextual comparison. However, since no universally established in-vitro %CV threshold has been defined in the literature for phantom-based DEXA studies, results are compared against values reported in prior cross-calibration studies using similar phantom-based protocols (Table 2). No published threshold was identified for soft-tissue composition %CV; therefore, the St-Thi and St-Att results are interpreted descriptively in relation to peer-reported values. Although Confidence intervals (95% CI) were calculated for all %CV estimates and p-value lower than 0.05 was considered statistically significant, however, for concise numerical reporting in the tables just significant results is marked.

In vitro short-term precision (IV-STP)

We performed a series of at least five consecutive scans, utilizing a combination of the lumbar phantom and five layers of Plexiglas, which is equivalent to a minimum of 10 centimeters of normal abdominal soft tissue thickness (part a). Following the completion of the workday, we repeated the process, performing at least five additional consecutive scans on the same combination of the lumbar phantom and Plexiglas layers (part b). The DEXA X-ray sources were exposed a minimum of 40 times during the intervals between parts a and b. Subsequently, we computed the %CV using the variation of the measured BMD values among parts a and b.

Repeatability test (RT)

To assess the repeatability of each DEXA system, the phantom was scanned three times consecutively without altering its position or configuration. This procedure was repeated 10 times, with each repetition involving a change in the thickness of the Plexiglas slabs above the lumbar phantom (ranging from two to 20 centimeters). In total, 60 scans were performed. The repeatability was quantified as the root mean square standard deviation, and used to calculate the %CV according to the ISCD's advanced precision calculation tool.

Stability across varying soft tissue thickness (St-Thi)

In this section, we evaluate the stability of the measured BMD in relation to changes in the thickness of Plexiglas placed above the lumbar phantom. For each specific Plexiglas thickness (ranging from two to 20 centimeters, in 2-centimeter increments), we performed three consecutive scans and registered the average of these consecutive scans as the calculated BMD value for that specific Plexiglas thickness. In total, for each system, we repeated the measurements 10 times, corresponding to 10 different Plexiglas thicknesses. We determined the range of thickness for which the device exhibited a stable response, which we refer to as the range of stable response area

(RSRA), for each system. Finally, we calculated the %CV over St-Thi for each system, considering variation of measured BMD values with Plexiglas thicknesses ranging from 10 to 20 centimeters.

Stability across varying soft tissue composition (St-Att)

From each of the six models of the DEXA systems examined in this study, one device was tested with a water phantom (equivalent to ~5% fat [8]), serving as a physical equivalent of soft tissue, in addition to Plexiglas (equivalent to ~40% fat [8]). The methodology was as follows: we positioned a water reservoir on the lumbar phantom and the water thickness varied from zero to 16 centimeters, with increments of two centimeters. For each water thickness, we conducted a minimum of three consecutive scans. The average BMD values obtained were plotted against the changes in water thickness. The mean BMD value for water thicknesses ranging from 10 to 16 centimeters was calculated as the measured BMD value with the water phantom. This BMD value was then compared to the equivalent BMD value obtained with the Plexiglas phantom. The variation in these BMD values is expressed as the %CV, taking into account the different soft tissue attenuation coefficients.

Accuracy test (AT)

Following the calculation of the RSRA for each DEXA system, we computed the average of the BMD values obtained within this region and compared it with the true BMD value of the lumbar phantom. Given that the BMD values reported on the phantom's sheet are attributed to the Norland company, the true BMD value was adjusted for each system in accordance with the conversion equation reported in [14-16]. The absolute percentage error between the averaged BMD over the RSRA and the adjusted true BMD was reported as the Accuracy. To facilitate a clear clinical comparison of the measurements across different DEXA systems, the standardized BMD values and the corresponding T-Score value, computed within the Caucasian or Non-Hispanic white race database of each system, is also provided.

RESULTS

Figure 3 includes Shewhart charts (a) to (g) related to the devices (a) Discovery, (b) Horizon-wi, (c) XR-800, (d) Lunar, (e) Osteo-Pro Max, (f) DEXXUM-T, and (g) Stratos DR. In these charts, the horizontal axis represents change in Plexiglass thickness from 10 to 20 centimeters, and the vertical axis represents the calculated percentage mean absolute error of the measured BMD from the lumbar phantom. Each diagram encompasses a set of 4 or 5 distinct curves, each of which corresponds to a specific device. In the diagrams of Figure 3, the gray lines are related to the $\pm 1.5\%$ error of the actual BMD of the phantom. The graph (a) corresponds to the discovery devices, and all the curves in this graph exhibit relatively stable behavior regarding changes in the thickness of the Plexiglass. The BMD value is less dependent on the thickness of the soft tissue. Additionally, the measurements are very close to the actual BMD value of the phantom. In each of the diagrams, the curve corresponds to Dis6 system (pink curve), which is characterized by its exceptional accuracy, is presented to facilitate more effective comparative analysis.

The Horizon-wi devices are represented by graph (b), where all the plotted curves demonstrate a relatively consistent response to variations in Plexiglass thickness. The BMD measurement shows less reliance on the thickness of soft tissue. Moreover, the recorded measurements closely align with the true BMD value of the phantom.

Graph (c) represents the XR-800 devices, and all the curves displayed in this graph demonstrate a relatively consistent response to variations in Plexiglass thickness. The measurements show less reliance on the thickness of soft tissue. Nevertheless, there is a greater disparity between the recorded measurements and the true BMD value of the phantom when compared to the Dis6 device (pink curve). This suggests a higher degree of error and lower accuracy in the measurements. However, the accuracy is within the previously published range (Table 2).

Graph (d) represents the lunar devices. The curves in this graph show consistent behavior when they come to changes in Plexiglass thickness and seems that they are not influenced by the thickness of soft tissue. There is a greater difference between the recorded measurements and the true BMD value of the phantom compared to the Dis6 (pink curve), also suggesting a higher degree of error but moderate accuracy.

The graph labeled (e) corresponds to the Osteo-Pro max devices. It is evident that these devices' response depends on variations in the thickness of soft tissue equivalent. Within the range of 10 to 20 centimeters, the response of the device is not consistent. The measurements deviate from the actual BMD value of the phantom, indicating a higher degree of error compared to the Dis6 (illustrated by the pink curve), resulting in diminished accuracy.

It can be observed in graph (f) corresponding to the Dexam-T devices that their response also exhibits dependence on the thickness of soft tissue and lacks stability. Furthermore, there is a significant difference between the measured values and the actual BMD value of the phantom. In comparison to the pink curve, they have a higher level of error and lower accuracy.

In graph (g) the Stratos devices exhibit a relatively stable response from 10 centimeters of Plexiglass thickness onwards. The measurements taken differ from the actual BMD value of the phantom, but the calculated error is relatively low, and they possess a comparable level of accuracy.

More specific numerical data pertaining to the aforementioned diagrams is provided in Tables 3 and 4. Table 3 provides an overview of the findings derived from measurements conducted on 22 densitometers. The results from identical model of DEXA systems are averaged to facilitate a more straightforward comparison. In assessing the second column of Table 3, which stems from IV-STP, it becomes evident that the highest %CV values are associated with the Horizon, Lunar and Osteo-Pro max systems (up to 1.49 %). Moreover, third column show the %CV in RT. So, all of these measurements fall within the range reported by [11] and [15] for similar phantom-based cross-calibration protocols (Table 2)

Furthermore, in the fourth column, which was results of St-Thi test, it is observed that the poorest %CV values can be attributed to the Osteo-Pro max (4.24%) and Dexam-T (6.65%) systems. Also, as illustrated in the fifth column, the %CV of St-Att for all systems is comparable, with the exception of the Dexam-T system, which stands at 3.87%. Additionally, the RSRA in Lunar, Osteo-Pro max, Dexam-T, and Stratos DR was found to be shorter than that in Horizon-wi, Discovery, and XR-800 (as indicated in the last column). This narrower RSRA corresponds to the systems that also exhibited accuracy and %CV values falling outside the reference ranges reported in the literature (>90% accuracy [10]; $\leq 1.9\%$ CV, ISCD 2023), as detailed below.

Table 4 delineates the average measured BMD $\pm$ SD of the lumbar phantom across different models. Additionally, the standard BMD, calculated from coefficients provided in prior studies, is also presented [14-16]. The corresponding T-score for each averaged BMD is derived from the Caucasian or Non-Hispanic white database, as provided by the system model software. It is

noteworthy that the documented BMD value of the phantom is 0.858 gr/cm². This BMD value corresponds to a T-score of -1.5 in the Caucasian dataset of the XR-800 system. Furthermore, the final column of Table 4 exhibits the average accuracy of each DEXA model. These accuracy values are extracted and computed from Shewhart charts. It is evident that the accuracies of the Osteo-Pro Max and DEXXUM-T models significantly deviate from those of other models. Notably, their accuracy values (89.85% and 86.50%, respectively) fall below the >90% benchmark previously reported in phantom-based literature [10,11], whereas the remaining systems meet or exceed this threshold, indicating comparatively high-performing status.

As a pilot, illustrative clinical case series, a small-scale cross-calibration comparison between the Osteo-Pro Max and DEXXUM-T systems and the XR-800 system from the Norland manufacturer was conducted with four female volunteers under an institutional ethics exemption obtained for this internal validation (see Declarations). Given the limited sample size (n=4), these findings should be interpreted as hypothesis-generating rather than conclusive evidence of clinical misclassification risk. The age of each volunteer was as follows: patient 1 (67 years old), patient 2 (26 years old), patient 3 (35 years old), and patient 4 (28 years old). Notably, the first volunteer was postmenopausal, while the remaining volunteers were healthy young Persian women of Caucasian ethnicity. The results of this analysis are presented in Table 5. Each volunteer underwent one scan with the Nrd9 DEXA system and subsequently repeated their scan on one of the aforementioned DEXA devices. As demonstrated, there is a clinically significant change in the measured BMD and their corresponding T-Scores between the two repeated scans. For instance, the BMD test of patient four conducted with the Nrd9 system resulted in a T-Score of -0.5, indicative of normal BMD. However, the same patient's BMD test with the Osts14 DEXA system yielded a T-Score of +1.3, and with the Osts16 DEXA system, a T-score of -1.8. These inter-system clinically notable variations were observed in a small validation subset and, while suggestive, should be interpreted with caution; larger patient cohorts are needed to confirm whether such variations translate into different clinical decisions.

DISCUSSION

This study aimed to compare different DEXA device brands across several key performance parameters using methods feasible for a one-day evaluation. The assessed metrics included precision, repeatability, stability, and accuracy of BMD measurements. While results for different models were compared in the previous section, here, we discuss these findings in light of prior research.

Table 2 summarizes precision and accuracy values reported in earlier studies for various DEXA systems. As shown in the second column, previous reports indicate that precision values of most DEXA systems are within 1.8% CV, except for the DEXXUM-T model. Our findings, presented in Table 3, are consistent with these values.

One notable observation in this study was the relatively low accuracy obtained from the Osteo-Pro Max and DEXXUM-T models, with average accuracies of 89.85% and 86.50%, respectively. These results suggest that these devices should be used cautiously. However, as indicated in Table 2, other studies have also reported accuracies around 86–88% in phantom-based experiments. Such deviations could potentially lead to misinterpretation of a patient's treatment response.

To further investigate, we compared the accuracy of the Osteo-Pro Max and DEXXUM-T systems with the XR-800 using BMD measurements in a small group of volunteers (n = 4). As shown in

How to cite this article:

Abbasi F, Faghihi R, Entezarmahdi SM, Gheisari F, Haghighatafshar M. Comparative analysis of quality control results of 22 dual-energy X-ray absorptiometry (DEXA) systems across six manufacturers using a comprehensive one-day protocol. Iran J Nucl Med. doi: 10.22034/irjnm.2026.130431.1734

Table 5, T-score values differed by more than one unit between systems in this limited sample. Given the small sample size, these findings should be considered preliminary and hypothesis-generating rather than conclusive evidence of clinical misclassification; a larger patient cohort is required to confirm whether this magnitude of difference is clinically relevant. Overall, our in vitro accuracy results highlight the need for cross-calibration between different DEXA systems before performing follow-up scans on the same patient using distinct devices.

Our results also showed that measured BMD values remained stable across most models when equivalent soft tissue thickness was varied. Laskey et al. conducted a similar St-Thi test and obtained comparable results using Lunar DPX and Hologic QDR-1000 systems, reporting BMD stability within 1% for tissue thicknesses between 15 and 26 cm, with a minor decrease at 30 cm [17]. Similarly, Blake et al. found negligible BMD variation with water thickness between 15–25 cm for the QDR-1000 [7].

However, this consistency was not observed in the Osteo-Pro Max and DEXXUM-T models in our study, where average %CVs for the St-Thi test were 4.24% and 6.65%, respectively. These results suggest that in these systems, measured BMD values may vary significantly with changes in soft tissue thickness, potentially affecting diagnostic reliability.

We also investigated the influence of soft tissue composition on device response. Our findings indicate that substituting Plexiglass with water as the soft tissue equivalent material did not significantly affect BMD values in most DEXA systems, except for the DEXXUM-T. Hansen et al. [9] previously reported that altering soft tissue composition (fat-to-lean ratio) without changing thickness had less than a 1% impact on BMD. The significant variation observed in the DEXXUM-T model when changing soft tissue composition appears anomalous and may reflect calibration issues.

Several limitations of this study should be acknowledged. First, the lumbar phantom used had rough vertebral boundaries, potentially introducing delineation errors during repeated scans. Second, only a single BMD reference value was used. As the relationship between measured and actual BMD is nonlinear in most systems, a more comprehensive QC approach should include multiple reference values with distinct, well-defined densities, such as aluminum slices. Third, some systems offer user-dependent processing options, such as patient thickness or null-area selection, which can influence calculated BMD values. Therefore, careful selection and validation of these options are essential. Additionally, the patient-based validation in this study was limited to a small cohort of four volunteers, which is insufficient to draw definitive conclusions regarding clinical misclassification or diagnostic unreliability between systems. Large-scale patient cross-validation studies with adequately powered sample sizes are needed to confirm these preliminary observations and to better understand variability in BMD and T-score values among different DEXA models. Moreover, manual adjustment of the region of interest (ROI) was occasionally required in cases where automatic edge detection was suboptimal. These manual corrections were performed by a single experienced nuclear medicine physician; however, the resulting operator-dependent variability was not formally quantified in this study. This represents a limitation of the present analysis, as inter-observer and intra-observer variability in manual ROI placement could potentially influence BMD measurements and, consequently, the reported precision and accuracy estimates.

A further limitation relates to acquisition-level technical parameters. On fan-beam DXA systems, tube potential (kVp) and tube current (mA) are automatically selected by the scanner's internal

algorithm based on measured tissue/phantom thickness and are not accessible or adjustable by the operator; these values are proprietary to each manufacturer and were not retrievable from the clinical acquisition software at the participating centers. While the analysis/scan mode (e.g., standard vs. thickness-tiered mode) was harmonized across all 22 systems as described in the Methods, we cannot rule out that residual inter-system heterogeneity in accuracy and precision — particularly the pronounced deviations observed in the Osteo-Pro Max and DEXxum-T systems — partly reflects manufacturer-specific hardware calibration or firmware-level technique-factor algorithms rather than center-specific software presets. Disentangling these two sources would require manufacturer-provided technical documentation or controlled bench testing under disclosed, fixed kVp/mA conditions, which was beyond the scope of this field-based comparative QC study. Future work involving direct collaboration with manufacturers to log or standardize technique factors is warranted to isolate the contribution of hardware versus software/preset variability.

Errors in DEXA measurements can also stem from inadequate equipment maintenance, insufficient operator training, or unfamiliarity with updated standards and guidelines. Proper calibration and adherence to manufacturer instructions are critical for accurate results. The variations observed in our study likely reflect differences in calibration methods or deficiencies in manufacturer-recommended QC procedures.

We do not claim superiority of one device over another in delivering accurate results, nor are we arguing that absolute accuracy is indispensable, although we believe it is highly desirable. The key concern is the unnecessary variability among instruments, which diminishes the clinical utility of bone densitometry. This issue could be mitigated if both users and manufacturers adopted standardized calibration and QC protocols using internationally recognized reference phantoms. The importance of such standardization has long been recognized. The American Association of Physicists in Medicine (AAPM) established a task force to develop recommendations for standard phantoms, aiming to harmonize DEXA performance assessment. We strongly advocate immediate implementation of standardized QC phantoms and protocols in all bone densitometry laboratories to enhance comparability and reliability.

Currently, the International Society for Clinical Densitometry (ISCD) does not provide unified guidelines for verifying calibration and accuracy across DEXA systems. Consequently, device calibration using manufacturer-supplied phantoms may not guarantee true accuracy. We recommend that ISCD develop clear and consistent guidelines for evaluating DEXA calibration and assume a supervisory role in ensuring their enforcement. Establishing such standards would be a crucial step toward global harmonization of BMD measurements.

Moreover, creating a reference phantom specifically designed for QC evaluation of DEXA systems would greatly improve inter-device comparability. If ISCD were to define manufacturing standards and implementation protocols for such reference phantoms, it could significantly advance quality assurance practices in bone densitometry.

Finally, implementing these measures should be complemented by the creation of a supervisory authority to oversee compliance and regulations. Assigning responsibility and establishing independent oversight would enhance confidence in the accuracy and reliability of BMD assessments across clinical centers.

How to cite this article:

Abbasi F, Faghihi R, Entezarmahdi SM, Gheisari F, Haghighatafshar M. Comparative analysis of quality control results of 22 dual-energy X-ray absorptiometry (DEXA) systems across six manufacturers using a comprehensive one-day protocol. Iran J Nucl Med. doi: 10.22034/irjnm.2026.130431.1734

CONCLUSION

In conclusion, in this study we evaluated 22 DEXA systems from six distinct manufacturers using a five-step QC protocol. These steps included: (1) in-vitro short-term precision, (2) repeatability, (3) stability across varying soft tissue thickness, (4) stability across varying soft tissue attenuation coefficient, and (5) accuracy. In this work we have demonstrated the limitations of ISCD recommended QC protocols for DEXA systems. Our findings revealed that without comprehensive quality control, the outcome of the BMD measurements of some vendors' systems fell outside the range previously reported in other publications, which may increase the risk of diagnostic error; however, confirming the clinical impact of this risk requires validation in larger patient cohorts. These results underscore the need for establishing a comprehensive standard QC protocol by international organizations. Furthermore, regular monitoring is advised to ensure that the system performance remains at an optimal clinical level.

Ethics approval and consent to participate

The main phantom-based quality control component of this study did not involve human or animal subjects and therefore did not require ethics committee review. For the supplementary cross-cultural analysis involving four volunteers (Table 5), a formal exemption permit was obtained from Shiraz University of Medical Sciences Research Committee (Exemption No. 1405.573, dated 2026/08/22), given that no additional radiation exposure beyond standard quality-assurance procedures was involved.

Acknowledgements

This work is a part of a Master thesis in medical radiation-nuclear engineering by Fatemeh Abbasi in Shiraz University. We extend our sincere appreciation to Mr. Arash Borhanazad and the team at Sina Parto Jam Company for their invaluable scientific and experimental contributions to this research. We also gratefully acknowledge all medical imaging centers that participated in this study. Their assistance was indispensable to the successful completion of our project.

REFERENCES

1. Bono CM, Einhorn TA. Overview of osteoporosis: pathophysiology and determinants of bone strength. *Eur Spine J*. 2003 Oct;12 Suppl 2(Suppl 2):S90-6.
2. El Maghraoui A, Koumba BA, Jroundi I, Achemlal L, Bezza A, Tazi MA. Epidemiology of hip fractures in 2002 in Rabat, Morocco. *Osteoporos Int*. 2005 Jun;16(6):597-602.
3. Kanis JA, Johnell O, Oden A, Jonsson B, De Laet C, Dawson A. Risk of hip fracture according to the World Health Organization criteria for osteopenia and osteoporosis. *Bone*. 2000 Nov;27(5):585-90.
4. Blake GM, Fogelman I. The role of DXA bone density scans in the diagnosis and treatment of osteoporosis. *Postgrad Med J*. 2007 Aug;83(982):509-17.
5. Benmalek A, Sabatier JP. Comparison and cross-calibration of DXA systems: ODX-240 and Sophos L-XRA versus Hologic QDR-4500, for spinal bone mineral measurement. Translation of a reference database. *Osteoporos Int*. 1998;8(6):570-7.
6. Genant HK, Guglielmi G, Jergas M. Bone densitometry and osteoporosis. Berlin, Heidelberg: Springer; 2012.

How to cite this article:

Abbasi F, Faghihi R, Entezarmahdi SM, Gheisari F, Haghighatafshar M. Comparative analysis of quality control results of 22 dual-energy X-ray absorptiometry (DEXA) systems across six manufacturers using a comprehensive one-day protocol. Iran J Nucl Med. doi: 10.22034/irjnm.2026.130431.1734

7. Blake GM, McKeeney DB, Chhaya SC, Ryan PJ, Fogelman I. Dual energy x-ray absorptiometry: the effects of beam hardening on bone density measurements. *Med Phys*. 1992 Mar-Apr;19(2):459-65.
8. Mazess RB, Barden HS, Bisek JP, Hanson J. Dual-energy x-ray absorptiometry for total-body and regional bone-mineral and soft-tissue composition. *Am J Clin Nutr*. 1990 Jun;51(6):1106-12.
9. Hansen MA, Hassager C, Overgaard K, Marslew U, Riis BJ, Christiansen C. Dual-energy x-ray absorptiometry: a precise method of measuring bone mineral density in the lumbar spine. *J Nucl Med*. 1990 Jul;31(7):1156-62.
10. Davidsson L. Dual energy X ray absorptiometry for bone mineral density and body composition assessment. Vienna: International Atomic Energy Agency; 2010.
11. Genant HK, Grampp S, Glüer CC, Faulkner KG, Jergas M, Engelke K, Hagiwara S, Van Kuijk C. Universal standardization for dual x-ray absorptiometry: patient and phantom cross-calibration results. *J Bone Miner Res*. 1994 Oct;9(10):1503-14.
12. Khan AA, Colquhoun A, Hanley DA, Jankowski LG, Josse RG, Kendler DL, Lentle B, Leslie WD, Lewiecki EM, O'Neill E, Robertson S, Syed ZA, Tanner SB, Webster D. Standards and guidelines for technologists performing central dual-energy X-ray absorptiometry. *J Clin Densitom*. 2007 Apr-Jun;10(2):189-95.
13. Abbasi F: Comparison of bone mineral densitometry reproducibility among different dxa units and precision assessment by In-Vitro method. MSc thesis. Shiraz University, Department of Radiation Engineering; 2024.
14. Steiger P. Standardization of measurements for assessing BMD by DXA. *Calcif Tissue Int*. 1995 Dec;57(6):469.
15. Kim JS. Cross calibration of dual energy X-ray absorptiometry equipment for diagnosis of osteoporosis: between domestic manufacturers and global manufacturers. *J Korean Soc Radiol*. 2018;12(7):833–44.
16. Nalda E, Mahadea KK, Demattei C, Kotzki PO, Pouget JP, Boudousq V. Assessment of the Stratos, a new pencil-beam bone densitometer: dosimetry, precision, and cross calibration. *J Clin Densitom*. 2011 Oct-Dec;14(4):395-406.
17. Laskey MA, Flaxman ME, Barber RW, Trafford S, Hayball MP, Lyttle KD, Crisp AJ, Compston JE. Comparative performance in vitro and in vivo of Lunar DPX and Hologic QDR-1000 dual energy X-ray absorptiometers. *Br J Radiol*. 1991 Nov;64(767):1023-9.

Table 1. The number and model of dual energy X-ray absorptiometry systems used in this study

Number of systems	Manufacturer	Model	Abbreviation
3	Hologic	Discovery	Dis
4	Hologic	Horizon-wi	Hrzn
3	Norland	XR-800	Nrd
3	GE Lunar	Lunar	Lnr
3	BM tech	Osteo-Pro max	Ostp
3	Osteo-Sys	Dexxum-T	Osts
3	Stratos	Stratos DR	Strs

Table 2. %CV and accuracy of DEXA systems from different manufacturer have been reported in previous studies

	% CV (Precision test)	Accuracy (%)	Ref.
Hrzn	0.3-1.6	88-97	[11]
Nrd	0.5-1.8	90-92	[11]
Lnr	0.3-1.1	86-97	[11]
Ostp	0.3-1.2	-	[15]
Osts	0.7-7.4	-	[15]
Strs	1.3	-	[16]

Table 3. Coefficient of variation from different studies, among different DEXA systems

Model (n)	%CV (IV-STP)	%CV (RT)	%CV (St-Thi)	%CV (St-Att)	RSRA (cm)
Hrzn (n=4)	1.22	1.12	1.43	0.90	0 - 20
Dis (n=3)	0.50	0.92	1.84	1.12	0 - 20
Nrd (n=3)	0.96	0.84	1.64	1.07	0 - 20
Lnr (n=3)	1.41	0.82	0.93	0.32	4 - 20
Ostp (n=3)	1.49	1.24	4.24*	1.80	8 - 20
Osts (n=3)	0.76	0.61	6.65*	3.87*	6 - 20
Strs (n=3)	1.01	0.99	1.32	0.92	4 - 20

n: Number of evaluated DEXA systems; CV: Coefficient of variation; IV-STP: In vitro - short term precision; St-Thi: stability across varying soft tissue thickness; St-Att: stability across varying soft tissue attenuation coefficient; RSRA: Range of the stable response area; *Statistically significant difference from Dis6 system (p-value < 0.05).

Table 4. Mean of Measured BMD and sBMD and T-score and accuracy among different DEXA systems

Model (n)	Measured BMD (gr/cm ²)	sBMD (gr/cm ²)	T-Score	Accuracy (%)
Hrzn (n=4)	0.860 ±0.010	0.925	-1.7	98.50
Dis (n=3)	0.872 ±0.014	0.938	-1.6	98.54
Nrd (n=3)	0.894 ±0.011	0.962	-1.1	95.67*
Lnr (n=3)	1.001 ±0.006	0.953	-1.2	95.63*
Ostp (n=3)	1.086 ±0.042	1.023	-0.7	89.85*
Osts (n=3)	1.093 ±0.022	1.053	-0.6	86.50*
Strs (n=3)	0.903 ±0.009	1.001	-1.0	94.70*

n: Number of evaluated DEXA systems; sBMD: Standardized BMD. Note: Phantom BMD: 0.858 (gr/cm²), (T-Score: -1.5); * Statistically significant change from the lumbar phantom BMD value (p-value < 0.05).

Table 5. Comparative BMD Values and T-scores from Norland XR-800 and Osteo-Pro max and Dexam-T DEXA systems for four volunteers

	Model	BMD (L1-L4)	T-score
Patient 1	XR-800 (Nrd9)	0.634	-2.8
	Osteo-Pro max (Ostp11)	0.710	-4.0
Patient 2	XR-800 (Nrd9)	0.994	-0.5
	Osteo-Pro max (Ostp12)	1.044	-1.3
Patient 3	XR-800 (Nrd9)	1.105	0.1
	Dexam-T (Osts14)	1.258	1.3
Patient 4	XR-800 (Nrd9)	0.993	-0.5
	Dexam-T (Osts14)	1.258	1.3
	XR-800 (Nrd9)	0.993	-0.5
	Dexam-T (Osts16)	0.883	-1.8

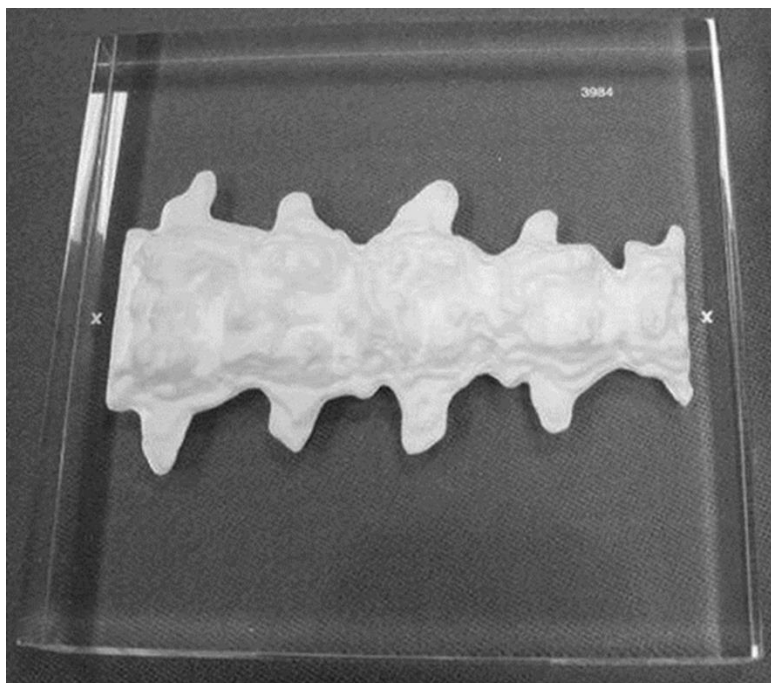


Figure 1. Top view of the Norland lumbar phantom



Figure 2. (Left) tissue-equivalent Plexiglas slabs placed above lumbar phantom, (Right) water tank placed above lumbar phantom

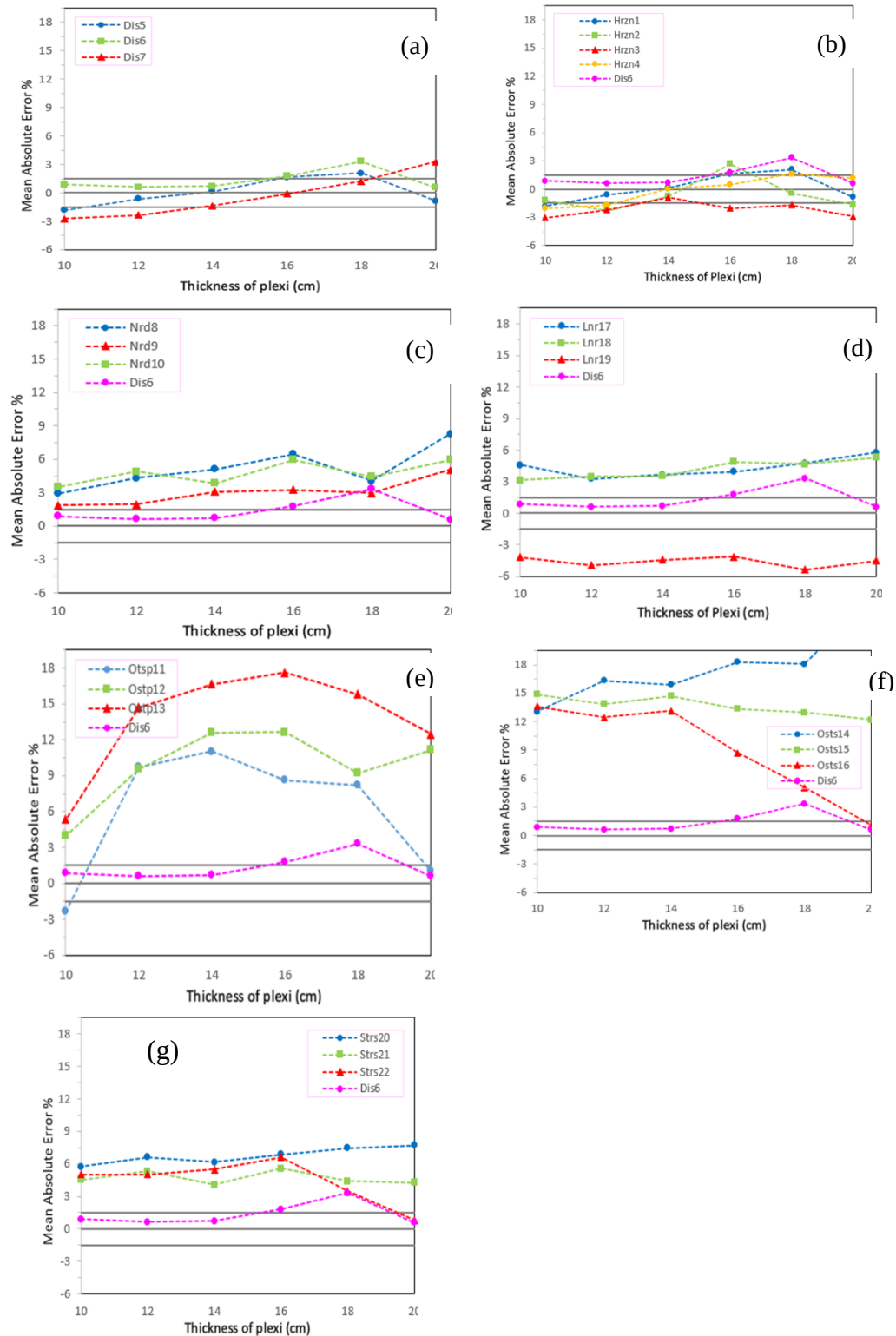


Figure 3. Shewhart charts: Evaluating the variation in percentage mean absolute error of measured BMD values across soft tissue phantom thickness. For comparative analysis, the curve corresponds to Dis 6 system is repeated in pink in all subsequent charts. (a) Discovery, (b) Horizon Wi, (c) XR-800, (d) Lunar, (e) Osteo-Pro Max, (f) Dexxum-T, and (g) Stratos DR