Evaluation of a new body composition phantom for quality control and cross-calibration of DXA devices

E. DIESSEL, T. FUERST, C. F. NJEH, F. TYLAVSKY, J. CAULEY, M. DOCKRELL, and H. K. GENANT

University of California San Francisco, San Francisco 94143-1349; Synarc, Inc., San Francisco, California 94105; University of Tennessee, Memphis, Tennessee 38103-4900; and University of Pittsburgh, Pittsburgh, Pennsylvania 15238-2887

Received 17 November 1999; accepted in final form 1 April 2000

Address for reprint requests and other correspondence: C. F. Njeh, Osteoporosis and Arthritis Research Group, Univ. of California San Francisco, 350 Parnassus Ave, Suite 908, San Francisco, CA 94143-1349 (E-mail: christopher.njeh@oarg.ucsf.edu).

Dual-energy X-ray absorptiometry; in vivo; percentage body fat; precision

Body composition measurements using dual-energy X-ray absorptiometry (DXA) devices are becoming more attractive in clinical practice, thanks to a greater awareness of their clinical utility and the reduced scanning time of newer systems. A body composition measurement requires scanning the total body and provides information about percent fat of the total body and body regions, lean soft tissue mass, and bone mineral mass (22). Long-term stability of DXA devices is an important prerequisite for reproducible body composition or bone density measurements in longitudinal studies. Quality control is usually performed for bone mineral density (BMD) measurements using spine or femur phantoms that represent defined BMD values. These phantoms are usually provided by the manufacturer for daily quality control purposes. In addition to quality control, these phantoms are used for cross-calibration among different DXA instruments from the same manufacturer in multicenter trials. However, only a few phantoms have been evaluated for quality control or cross-calibration of body composition measurements such as soft tissue mass (lean and fat) or percent fat (13, 15).

Paton et al. (23) found significant in vivo differences in fat and lean tissue readings when comparing identical models of DXA devices from the same manufacturer (DPX models, Lunar, Madison, WI) at different locations. The devices were comparable according to BMD readings and passed all quality control tests but were not comparable according to lean and fat tissue readings. In other in vivo studies, differences in body composition results were found between Hologic pencil and fan beam devices (1, 3) and between various scan modes (26). Differences in fat and lean tissue results are expected in comparisons of DXA devices from different manufacturers, e.g., Lunar and Hologic (20, 25), because significant differences are known to exist with regard to BMD readings (7, 28). Reasons for different readings between manufacturers can be due to differences in calibration of the devices, in software algorithms (e.g., bone detection, assumptions regarding distribution of soft tissue above bone), and in hardware (e.g., different X-ray energy spectra). Tothill et al. (27) found differences up to 13 ± 4% for percent fat measurements in the trunk using Norland and Hologic devices.

Designed primarily for validation studies of body composition measurements, phantoms have been man-
ufactured with variable amounts of water or meat to represent fat-free tissue, lard or ethanol to represent fat, and aluminum or bone to represent bone (10, 11, 16, 24). Jensen et al. (13) developed two phantoms that were used for quality control over a period of 1 yr: one consisted of ground meat blocks, and the second consisted of human bone with water and lard. The materials were kept frozen when not in use.

An anthropomorphic-like phantom was designed by Nord for Norland (21). It consists of a simplified skeleton of aluminum sheets, with sheets of acrylic and thin vinyl plastic representing soft tissue. Thickness can be altered, and the effective fat proportion can vary from 22 to 40% fat. Tothill et al. (28) tested the phantom on three DXA devices, one each from Lunar, Hologic, and Norland, and found good results compared with in vivo data; however, this phantom is not commercially available.

Two new body composition phantoms have recently been introduced by Hologic and Lunar. The Hologic phantom consists of six rectangular blocks of high-density polyethylene and weighs 40 kg. An aluminum skeleton is embedded in the base layer to simulate the bones of the head, arms, spine, pelvis, and thighs. An additional sheet of polyvinylchioride (PVC) is embedded, simulating lean tissue. The Hologic phantom was scanned for a period of 1 yr and proved useful for long-term quality control of DXA scanners used in body composition measurements (15). Short-term precision was very good (0.2, 0.5, and 0.3% for lean mass, fat mass, and percent fat, respectively) using a QDR-4500 scanner.

No study has been conducted to test the Lunar variable composition phantom (VCP). The purpose of this study was to evaluate the new Lunar body composition phantom's usefulness in quality control and to compare the results with in vivo data. The phantom tested here is similar to the commercial version that is currently available from Bio-Imaging (West Trenton, NJ). Slight differences exist between the commercially available phantom, the phantom tested for the manufacturer-provided nominal percent fat values, and the aluminum object that is used to simulate the bones of the head.

METHODS

Body composition phantom. The VCP (Lunar) consists of four acrylic blocks (20 × 28 cm total dimension, ~15 cm thick, and 10 kg total wt), two thin PVC sheets, and four thin sheets of vinyl. The blocks and sheets are used in various combinations to simulate five different soft tissue compositions. The X-ray attenuation characteristics of the acrylic blocks are similar to adipose tissue, thus simulating fatlike tissue. By adding the thin PVC sheets, which have very lean mass attenuation characteristics, the fat proportion is decreased. The minimum fat proportion (13%) can be obtained by adding all six PVC and vinyl sheets, whereas the maximum fat proportion (43%) can be obtained by using only one PVC sheet.

DXA measurements. The VCP was scanned on eight different DXA devices, including the Hologic QDR-4500 and QDR-1500 and the Lunar Expert, DPX-IQ, and DPX-L. The QDR-4500 was used at four different centers, including Memphis, TN, Pittsburgh, PA, Gaithersburg, MD, and San Francisco, CA. According to Lunar recommendations, we measured the VCP with an aluminum block (the Lunar aluminum spine phantom) to represent the skull, as was required by the DPX software. In contrast, the Hologic software does not require a skull to be present, and the VCP was scanned without the additional aluminum block. The “lean” portion of the skull was, therefore, not included in the analyses of the total body scans; however, the same percent fat results were used when selecting a small region of interest (subregion) within the phantom for composition measurements. The phantom was scanned in the total body mode on all devices. For the three Lunar devices (DPX-IQ, DPX-L, and Expert), a medium scan mode was selected, according to Lunar recommendations. For the Hologic devices, the standard scan mode for total body scans was chosen. Duplicate measurements of the VCP phantom were performed for each of the five configurations on all DXA devices. In this paper, we present the average values of the duplicate scans for each configuration. In this study, the scan analysis differed slightly for each device. For DPX devices (pencil beam), the standard whole body analysis was used, as was recommended by Lunar. For the QDR-4500 (fan beam), we analyzed only a subregion within the VCP to avoid the inclusion of X-rays that did not penetrate the entire thickness and all layers of material due to the diverging fan beam geometry. Unfortunately, the Lunar Expert (fan beam) software (version 1.72) that we used did not allow us to select a region of interest within the phantom area; thus the phantom was analyzed as a total body including the aluminum skull phantom.

Subjects. A total of 130 healthy female and male volunteers, with a wide range of weight and body mass indexes, were recruited. This study was approved by the ethical committee of each center, and subjects gave informed consent. Each volunteer underwent whole body scans on two different DXA devices (Table 1). Scans were performed in two different centers that participated in the health, aging, and body composition study (health ABC): Memphis, TN, and Pittsburgh, PA. Measurements were also carried out in San Francisco, CA; this center, however, was not a recruitment site for the health ABC study but did participate in the health ABC validation study.

Statistical analysis. Patient and phantom percent fat values measured on the different scanners were compared using linear regression. Midterm precision was determined for the Hologic QDR-4500 and the Lunar DPX-IQ scanners in San Francisco, CA, by scanning three configurations (lowest, medium, and highest percent fat values) on 15 different days within a 3-mo period. Only limited data for determining

Table 1. Healthy subjects scanned on various DXA scanners from different centers

<table>
<thead>
<tr>
<th>No. of Patients</th>
<th>DXA Scanners (City)</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>DPX-IQ(SF), QDR-4500(SF)</td>
</tr>
<tr>
<td>75</td>
<td>QDR-1500(P), QDR-4500(P)</td>
</tr>
<tr>
<td>75</td>
<td>QDR-1500(M), QDR-4500(M)</td>
</tr>
<tr>
<td>10</td>
<td>QDR-4500(P), QDR-4500(M)</td>
</tr>
</tbody>
</table>

Number of subjects scanned on different dual-energy X-ray absorptiometry (DXA) scanners in centers participating in the Health, Aging, and Body Composition study in San Francisco, CA (SF), Pittsburgh, PA (P), and Memphis, TN (M). DPX-IQ, DXA scanner model from Lunar (Madison, WI); QDR-1500 and QDR-4500, DXA models from Hologic.
RESULTS

The VCP covered a wide clinical range, with nearly 90% of the percent fat measurements in the in vivo studies falling within the range of the nominal percent fat values of the VCP (13–43% fat). Mean midterm precision of the percent fat values given as standard deviation (SD) was 0.6% for the QDR-4500 and 0.8% for the DPX-IQ. Mean short-term precision values (as absolute SD), calculated using limited data, were 0.7, 0.3, and 0.6% for QDR-1500, DPX-L, and Expert, respectively. The SD values for the different percent fat values were similar for each device. However, when referring to the precision as the coefficient of variation (CV = SD/mean percent fat for each configuration), measurements with high-percent fat range (43%) were more precise than those in the low-percent fat range (13%; CV = 1.5 vs. 5%, respectively).

Table 2 shows the deviation from the nominal percent fat value for the different configurations and the different devices. The nominal value established by Lunar only pertains to the DPX devices. DPX-IQ agreed best with the nominal values of the VCP, whereas we found differences from the nominal values for all other scanners (up to 7% for Lunar Expert). This suggests there are differences in the fat calibration of these scanners. The percent fat values measured on the four QDR-4500 devices were generally in agreement with each other. The largest difference (−1.5%) in percent fat was found between the QDR scanners in Memphis, TN, and Pittsburgh, PA (Table 3).

Results of the comparisons between phantom and in vivo data are shown in Figs. 1–4. There was a highly significant correlation (r² = 0.96–0.99) between the different scanners with regard to both in vivo and in vitro data. The measurements from Memphis, TN, showed the best agreement between the respective regression lines through patient and phantom data when comparing the Hologic QDR-1500 (pencil beam devices) with the QDR-4500 (fan beam devices) (Fig. 1). Both the slope and intercept of the patient data were not significantly different from the phantom data; however, for the Pittsburgh, PA, center, both the slope and intercept of patient data (0.791 and 2.67, respectively) were statistically different (P < 0.001) from phantom data (slope 0.868, intercept 0.79; Fig. 2). However, when patient data were forced through the origin, the gradient (0.861) was similar to that of the phantom. It is worth noting the similar behavior of the phantoms at Memphis, TN, and Pittsburgh, PA. Figure 3 shows the comparison of two QDR-4500 scanners. The slope of 1.0 is the same for both the patient and phantom regression lines, with a nonsignificant intercept. The Lunar DPX-IQ (pencil beam) tends to measure higher percent fat values in vivo than the QDR-4500 (Fig. 4), and this tendency was reproduced with the VCP measurements. However, the regression lines show that the slope and intercept of the patient data are significantly (P < 0.05) different from those of the phantom data (Fig. 4).

DISCUSSION

Phantoms play an important role in bone densitometry. As a device used for taking quantitative measure-

Table 3. Percent fat values of VCP soft tissue compositions measured on the Hologic QDR-4500 scanner in San Francisco and deviations for three other Hologic QDR-4500 scanners

<table>
<thead>
<tr>
<th>VCP Soft Tissue Compositions</th>
<th>San Francisco, CA</th>
<th>Gaithersburg, MD</th>
<th>Memphis, TN</th>
<th>Pittsburgh, PA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal %fat</td>
<td>40.1%</td>
<td>0.0%</td>
<td>−0.9%</td>
<td>0.3%</td>
</tr>
<tr>
<td>QDR-4500 (Hologic, SF)</td>
<td>36.8%</td>
<td>0.0%</td>
<td>−1.0%</td>
<td>−0.4%</td>
</tr>
<tr>
<td>QDR-1500 (Hologic, P)</td>
<td>27.0%</td>
<td>−0.3%</td>
<td>−1.3%</td>
<td>0.2%</td>
</tr>
<tr>
<td>QDR-1500 (Hologic, M)</td>
<td>20.2%</td>
<td>0.0%</td>
<td>−1.2%</td>
<td>−0.4%</td>
</tr>
<tr>
<td>Expert (Lunar)</td>
<td>13.4%</td>
<td>−0.2%</td>
<td>−0.7%</td>
<td>0.2%</td>
</tr>
</tbody>
</table>

Values are in absolute percent fat values. Mean percent fat values of duplicate measurements were used for each calculation.
ments of bone mineral or fat percentage, a DXA scanner must be precise, accurate, and stable over time. Phantoms are often used to evaluate these characteristics. In practice, phantoms can be used for two purposes: 1) cross-sectional cross-calibration, in which phantoms are used to compare the calibration of a group of scanners, with the goal of pooling the data from these different devices, and 2) longitudinal quality control, in which serial measurements are used to monitor scanner performance, with the goal of detecting instability in calibration that might affect the interpretation of longitudinal measurements.

The success of a phantom to achieve these goals rests heavily on its design. A phantom can be designed to test fundamental properties of scanner performance or to test scanner performance in aggregate. In the former case, phantoms are generally of simple design and several are possibly required to test all relevant characteristics. In the latter case, the phantom design becomes more complicated as it tries to simulate, as close

![Graph 1](image1.png)

**Fig. 1.** Percent fat values of 75 patients scanned on Hologic QDR-1500 and QDR-4500 scanners in Memphis, TN. Squares, patient data (% fat_{QDR-4500} = 1.30 + 0.832 \times \text{fat}_{QDR-1500}). Diamonds, phantom data (% fat_{QDR-4500} = 0.64 + 0.823 \times \text{fat}_{QDR-1500}).

![Graph 2](image2.png)

**Fig. 2.** Percent fat values of 75 patients scanned on Hologic QDR-1500 and QDR-4500 scanners in Pittsburgh, PA. Patient data (squares) was determined as % fat_{QDR-4500} = 2.67 + 0.791 \times \text{fat}_{QDR-1500}; phantom data (diamonds) as % fat_{QDR-4500} = 0.79 + 0.868 \times \text{fat}_{QDR-1500}.

![Graph 3](image3.png)

**Fig. 3.** Percent fat values of 75 patients scanned on Hologic QDR-4500 devices, one each from Memphis, TN, and Pittsburgh, PA. Patient data (squares): % fat_{Pittsburgh} = -0.66 + 1.019 \times \text{fat}_{Memphis}; Phantom data (diamonds): % fat_{Pittsburgh} = 0.92 + 1.003 \times \text{fat}_{Memphis}.

![Graph 4](image4.png)

**Fig. 4.** Percent fat values of 45 patients scanned on Lunar DPX-IQ and Hologic QDR-4500 in San Francisco. Patient data (squares) determined as % fat_{QDR-4500} = 3.68 + 0.807 \times \text{fat}_{DPX-IQ}; phantom data (diamonds) as % fat_{QDR-4500} = -2.35 + 0.966 \times \text{fat}_{DPX-IQ}.
as possible, the in vivo situation. A phantom for whole body scanning by DXA should have the following characteristics: 1) It is practical for frequent scanning and shipment among study centers. 2) It can be scanned and analyzed using standard manufacturer-provided protocols. 3) The phantom is able to simulate the range of masses, tissue thicknesses, and tissue compositions that are characteristic of the clinical population of interest. 4) The phantom includes bone segments that can be used to test mineral calibration; 5) is composed of materials that are inexpensive, machinable, temporarily stable, and closely matched to the physical characteristics of human tissues; and 6) is sensitive to scanner calibration differences or component malfunction and aging that may influence in vivo results.

A phantom that achieves all of these goals is difficult to design. A large and complex phantom that closely mimics the in vivo situation will probably not be easy to use. Moreover, just as all people are different, a phantom can be thought of as a single person and may not be representative of all people, or even the average person. Therefore, it becomes important to understand the behavior of the phantom, and, in particular, how it relates to measurements in people and what limitations there are in extrapolating from in vitro to in vivo situations. The VCP is a reasonable compromise of these goals. It is relatively small and lightweight, making it easy to use. The five variable configurations allow evaluation of a wide range of body fat percentages; however, its small mass (~10 kg) and nonanthropomorphic design may limit its ability to fully test scanner performance.

Our results showed that the VCP can be used to investigate body fat percent measurements made with various DXA devices and that the VCP could be scanned on all instruments. Precision data collected in this study showed that the VCP has the potential for use as part of a longitudinal quality control program to ensure the stability of a DXA device. Absolute precision error was low, ranging from 0.6 to 0.8%. The CV of the VCP scans (1.5–5%, depending on percent fat value) was found to be poorer than the precision of spine phantom scans (CV = 0.4–0.7%) for the quality control of bone mineral measurements (4, 6, 17). Nevertheless, the absolute precision (SD = 0.6–0.8%) compares well against typical, clinically relevant changes in body fat percentage (5–10%). Whether the VCP is able to detect clinically relevant drifts in soft tissue evaluation due to device instability depends not only on its precision but also on its sensitivity to scanner drift. Further research is required to answer this question. In addition, other questions remain to be addressed. For example, scanning protocols that recommend how often and which configurations of the VCP should be scanned for quality control purposes need to be established. It may be sufficient to scan the phantom less frequently (e.g., 2–3 times per week) or to measure only a subset of configurations. Our results showed a linear relationship between the different percent fat values of the different configurations for all devices in this study, suggesting it may not be necessary to scan all configurations. Thus daily quality control could be limited to the highest and lowest percent fat configurations (43 and 13%, respectively). Another issue is the correction of in vivo body composition data using the VCP scans. This issue was not addressed in this study and requires further investigation.

The CV of the VCP for percent fat values (1.5–5%) is comparable or worse than that found in vivo. In vivo CV for percent fat measurements for the total body varied between 0.7 and 1.8% for the QDR-1000 (12, 25, 29) and between 1.4 and 3.0% for the DPX devices (5, 14, 18). One reason for this poorer precision can be attributed to the size of the phantom, which represents only a small fraction of the human body. With larger masses, more precise measurements (expressed as percentage) can be obtained. The Hologic body composition phantom, which is much larger, reported short-term precision of 0.3% for percent fat (15). Tothill et al. (29) also found a slightly worse precision for percent fat values in vitro than in vivo using the whole body phantom from Norland.

We also investigated the performance of the VCP for cross-calibration of DXA scanners. There was significant agreement between VCP measurements and patient data for the Hologic QDR devices that were used in the study. The tendency of the QDR-1500, for example, to measure a higher percent fat value than the QDR-4500 in the same patient could be simulated by the VCP. Similar, but not identical, regression lines were achieved. This finding suggests that the VCP may be useful for cross-sectional cross-calibration between QDR scanners and, perhaps more generally, for any scanners of the same make. However, the VCP may be less effective for cross-calibration between Lunar and Hologic scanners. In our comparison of scans from the DPX-IQ and QDR-4500 scanners, the VCP showed a different relationship between the two scanners than that seen in the patients. According to the data presented here, corrections or comparisons based on the phantom data alone would be incorrect. Further investigation is necessary to evaluate the use of the VCP for cross-sectional cross-calibration of DXA devices that were not involved in this study.

Cross-calibration must also be performed when a newer model is replacing an old scanner or when device parts that have influence on the calibration of the scanner (e.g., X-ray tube) are repaired or replaced. If longitudinal measurements will be continued on the new scanner or repaired scanner, a very precise comparison of calibration is required to ensure that the change in percent body fat evaluated for an individual is accurately assessed. The VCP may not represent the in vivo differences closely enough to allow a proper cross-calibration. In bone densitometry, it is known that phantom measurements alone may not be sufficient for scanner upgrade and/or replacement cross-calibration between devices and an in vivo cross-calibration experiment may be required (2). However, in vivo cross-calibration is difficult to perform, and phantoms may remain the only alternatives, in spite of their limitations (4, 9). We recommend against using the
VCP for scanner upgrade and/or replacement cross-calibration and suggest, instead, that in vivo comparisons be made in an adequate number of volunteers (15 to 20 subjects to cover a wide percent fat range).

The main limitations of the VCP are that it is not anthropomorphic and has a small mass. However, a larger-sized phantom that simulates the trunk, arms, legs, and so forth, would be significantly heavier and less user-friendly. A large, bulky phantom may never be accepted for routine clinical use. Thus the small size of the VCP has certain advantages.

Another limitation of the VCP is its lack of bone-equivalent. The VCP does not contain any bone-simulating components, except the aluminum spine block that simulates the skull. The skeleton plays an important role in a body composition phantom for three reasons. First, a skeleton would allow cross-calibration and longitudinal monitoring of bone mineral results, as well as percent fat. Whereas this study did not specifically test this, it seems that the small amount of bone-equivalent material in the VCP would not be adequate for evaluation of the bone mineral calibration of the whole body scan. A larger, anthropomorphic skeleton is likely required. Second, it is known that different DXA manufacturers use different image segmentation or edge detection techniques to separate bone from soft tissue (19). Differences in the delineation of bone not only affect bone mineral results but also affect the soft tissue quantification and percent fat measurement. Differences in bone detection are expected in areas of low BMD such as the ribs, fingers, and parts of the pelvis; however, because the VCP has no skeleton, these differences cannot be evaluated. The third reason that a skeleton is important in a body composition phantom is that DXA cannot directly measure percent fat in regions of the scan that contain bone mineral. Instead, the DXA software must interpolate the percent fat from measurements made in adjacent regions. In a typical whole body scan of a human being, fat and lean proportion can be determined in only ~60% of the body area that is bone free. The fat proportion in the remaining body area, which represents the skeleton, must be interpolated. Each manufacturer can be expected to use different fat distribution models for the purpose of interpolation of fat and lean tissue over bone (22, 27). These estimations can be expected to differ between systems and cannot be tested with the VCP.

These limitations may explain why the VCP performed better when comparing the results from scanners of the same make, whereas intermanufacturer comparisons showed worse agreement with the in vivo data. Scanners from the same manufacturer use the same algorithms and assumptions, and a phantom does not need to evaluate the effects of these algorithms. However, when comparing scanners from different manufacturers, the phantom should evaluate the impact that different algorithms will have on percent fat determination. It must be kept in mind that the human body is an extremely complicated object to simulate and any reasonably priced phantom can only take a small step toward a more accurate representation of the body. A phantom that fully tests DXA scanner performance may be neither achievable nor practical. Thus it is important to understand the strengths and weaknesses of simple phantoms such as the VCP.

In conclusion, the VCP is a simple, easy-to-use tool that shows good potential for use in longitudinal quality control of percent body fat measurements by DXA devices. The VCP measured similar percent fat values between DXA scanners of the same make and model. Moreover, when comparing DXA scanners of different models from the same manufacturer, the VCP detected differences in percent fat values that were similar to those found by in vivo comparison. This suggests that the VCP is capable of accurately measuring the fat calibration of DXA devices and may be suitable for cross-sectional cross-calibration between scanners from the same manufacturer. For the comparison of DXA scanners from different manufacturers, in vivo cross-calibration remains the only accurate method. Further studies are necessary to determine whether VCP data can be used for the correction of changes in body composition measurements due to technical failure. It is necessary to determine nominal calibration data for each make and model of DXA devices and to further investigate how to use the VCP for cross-calibration between different DXA models.

We thank the Health, Aging, and Body Composition (health ABC) study group, in particular, T. Harris from the National Institute on Aging. We are also very grateful to G. Binzer for collecting in vivo and phantom data; to V. Vandenberg for proofreading this manuscript; and to Dr. Shoujun Zhao of University of California San Francisco for assistance with statistical analysis.

Health ABC was supported by grants from the National Institute on Aging.

REFERENCES


8. Glauc CC, Blake G, Lu Y, Blunt BA, Jergas M, and Genant HK. Accurate assessment of precision errors: how to measure the...