Ultrasound in Med. & Biol., Vol. ■, No. ■, pp. 1–20, 2014

0301-5629/$ - see front matter

http://dx.doi.org/10.1016/j.ultrasmedbio.2014.08.017

● *Original Contribution*

A NOVEL ULTRASOUND METHODOLOGY FOR ESTIMATING SPINE MINERAL DENSITY

Francesco Conversano,* Roberto Franchini,* Antonio Greco,† Giulia Soloperto,* Fernanda Chiriacò,* Ernesto Casciaro,* Matteo Aventaggiato,† Maria Daniela Renna,* Paola Pisani,* Marco Di Paola,* Antonella Grimaldi,‡ Laura Quarta,‡ Eugenio Quarta,‡ Maurizio Muratore,‡ Pascal Laugier,§ and Sergio Casciaro*

*National Research Council, Institute of Clinical Physiology, Lecce, Italy; †Echolight srl, Lecce, Italy; ‡O.U. of Rheumatology, "Galateo" Hospital, San Cesario di Lecce, ASL-LE, Lecce, Italy; and §Laboratoire d'Imagerie Biomédicale, Sorbonne Universités, UPMC 06, INSERM, CNRS, Paris, France

(*Received* 7 *March* 2014; *revised* 14 *August* 2014; *in final form* 20 *August* 2014)

Abstract—We investigated the possible clinical feasibility and accuracy of an innovative ultrasound (US) method for diagnosis of osteoporosis diagnosis of the spine. A total of 342 female patients (aged 51–60 y) underwent spinal dual X-ray absorptiometry and abdominal echographic scanning of the lumbar spine. Recruited patients were subdivided into a reference database used for US spectral model construction and a study population for repeatability and accuracy evaluation. US images and radiofrequency signals were analyzed via a new fully automatic algorithm that performed a series of spectral and statistical analyses, providing a novel diagnostic parameter called the *osteoporosis score* (O.S.). If dual X-ray absorptiometry is assumed to be the gold standard reference, the accuracy of O.S.-based diagnoses was 91.1%, with $k = 0.859$ ($p < 0.0001$). Significant correlations were also found between O.S.-estimated bone mineral densities and corresponding dual X-ray absorptiometry values, with r^2 values up to 0.73 and a root mean square error of 6.3%–9.3%. The results obtained suggest that the proposed method has the potential for future routine application in US-based diagnosis of osteoporosis. (E-mail: sergio.casciaro@cnr.it) © **2014 World Federation for Ultrasound in Medicine & Biology.**

Key Words: **Osteoporosis diagnosis, Quantitative ultrasound, Bone mineral density measurement, Lumbar spine, Bone densitometry, Radiofrequency signal analysis.**

INTRODUCTION

Osteoporosis is a systemic skeletal disease characterized by low bone mass and micro-architectural deterioration of bone tissue, with a consequent increase in bone fragility and fracture risk (Liu et al. 2011). Unfortunately, this pathology is often still underdiagnosed (Curtis and Safford 2012; van den Bergh et al. 2012), resulting in a high incidence of osteoporotic fractures worldwide (Baim and Leslie 2012)—a serious problem for the elderly population and an enormous cost for national health care systems (Byberg et al. 2012; Pike et al. 2010; Sambrook and Cooper 2006). Therefore, introducing novel methods for anticipating and improving osteoporosis diagnosis and patient management is of great importance.

Currently, dual X-ray absorptiometry (DXA) is the most widely used method for diagnosing osteoporosis and is considered the "gold standard" reference for measuring bone mineral density (BMD) (Baim and Leslie 2012; Link 2012; Nayak et al. 2006; Pais et al. 2010; Schnitzer et al. 2012). BMD measurements are the basis of the operational definition of osteoporosis provided by the World Health Organization (WHO): osteoporosis is diagnosed when the hip or spine BMD is ≥2.5 standard deviations (SD) lower than the young adult mean (Kanis 1994). Hip and spine are the reference anatomic sites because it has been found that BMD measurements performed at these sites are the most accurate general predictors of osteoporotic fractures (Lewiecki 2010), together with the related high degree of health impairment and worse quality of life in the case of fractures at these anatomic sites. In particular, hip BMD is

Address correspondence to: Sergio Casciaro, Consiglio Nazionale delle Ricerche, Istituto di Fisiologia Clinica (CNR-IFC), c/o Campus Ecotekne (Ed. A7), via per Monteroni, 73100 Lecce, Italy. E-mail: sergio.casciaro@cnr.it

a stronger predictor of hip fracture than BMD measured at other sites (Cummings et al. 1993, 2006), and spine BMD is the preferred choice for treatment monitoring because of its high sensitivity to BMD variations (Lewiecki 2010).

Dual X-ray absorptiometry measurements cannot, however, be employed for population mass screenings because of several intrinsic limitations, related mainly to exposure to ionizing radiation with associated risks (Brambilla et al. 2013; Picano and Matucci-Cerinic 2011; Picano and Vano 2012; Semelka et al. 2012), high cost and the need for dedicated structures with certified operators. This has led to increasing interest in the investigation of quantitative ultrasound (QUS) methods for osteoporosis screening purposes (Breban et al. 2010; Nayak et al. 2006; Paggiosi et al. 2012; Pais et al. 2010; Schnitzer et al. 2012; Trimpou et al. 2010). Proposed QUS methods have several potential advantages over DXA (absence of ionizing radiation, portable machines, lower cost), but as yet there is no widespread consensus regarding their accuracy in identifying osteoporotic patients. In fact, commercially available QUS devices currently work only on peripheral sites (*e.g.*, calcaneus), and several studies investigating the correlations between QUS parameters and DXA-measured BMDs at the reference sites obtained contradictory results (Breban et al. 2010; Dane et al. 2008; El Maghraoui et al. 2009; Iida et al. 2010; Kwok et al. 2012; Liu et al. 2012; Moayyeri et al. 2012; Schnitzer et al. 2012; Stewart et al. 2006; Trimpou et al. 2010).

The most common QUS devices employ through-transmission measurements to provide parameters such as broadband ultrasound (US) attenuation, speed of sound and stiffness index. Recently, some experimental studies have reported the potential of ultrasonic backscattering as a new method for diagnosing osteoporosis, exploring the possible usefulness of parameters such as backscatter coefficient (Wear 2008; Wear et al. 2012), apparent integrated backscatter (AIB) (Hoffmeister et al. 2008; Jiang et al. 2014; Karjalainen et al. 2009, 2012), frequency slope of apparent backscatter and time slope of apparent backscatter (Hoffmeister et al. 2008), spectral centroid shift (Garra et al. 2009; Jiang et al. 2014), broadband ultrasound backscatter (Karjalainen et al. 2009; Padilla et al. 2008; Roux et al. 2001), integrated reflection coefficient (Karjalainen et al. 2009, 2012), mean of backscatter difference spectrum and slope of backscatter difference spectrum (Hoffmeister et al. 2012). The overall conclusions that can be drawn from the reported articles are that US backscatter parameters, mostly measured *in vitro* on excised human bone samples, have appreciable correlations with BMD, and experimental data often support the idea that backscatter measurements may also provide an assessment of bone micro-architecture. However, despite preliminary encouraging *in vivo* results reported by a few pilot studies (Garra et al. 2009; Karjalainen et al. 2012; Roux et al. 2001), the backscatter approach has still remained at an early stage of research and generally suffers from the lack of appropriate clinical validation, because the only available *in vivo* study involving a significant number of patients is the very recent work by Jiang et al. (2014), which included 1011 Asian patients of both genders, aged 21–80 y, and reported correlation coefficients up to 0.75 between US backscatter parameters measured at the calcaneus and DXA-measured BMD of the central reference sites.

In this context, the current official position of the International Society for Clinical Densitometry (ISCD) regarding QUS is that the only validated skeletal site for the clinical use of QUS in osteoporosis management is the heel; validated heel QUS devices predict fragility fractures in patients $>$ age 65 y and, in conjunction with clinical risk factors, can be used to identify a population at very low fracture probability in which no further diagnostic evaluation may be necessary (Krieg et al. 2008; ISCD 2013). However, ISCD also specifies that DXA measurements at the spine and femur are the preferred choice for therapeutic decisions and should be used in place of QUS if possible, and in particular, QUS cannot be used for therapeutic monitoring purposes (ISCD 2013).

As a consequence, diagnosis and management of osteoporosis in clinical routine are currently based on the evaluation of DXA measurement outcomes and presence of additional risk factors (Edwards et al. 2013; Ferrari et al. 2012), resulting in the reported evidence of underdiagnosis and undertreatment (Curtis and Safford 2012; van den Bergh et al. 2012).

A possible way to improve this situation could be the development of US-based approaches for non-ionizing BMD measurements at the reference anatomic sites. In fact, although the correlation between peripheral QUS parameters and DXA-measured central BMD (*i.e.*, spinal or femoral) is typically poor, site-matched correlations between DXA-measured BMD and corresponding QUS estimates are generally much stronger (Barkmann et al. 2010). Therefore, an improved diagnostic outcome could be expected from US measurements on the reference central sites.

On the basis of these considerations, proximal femur has become the target of several recent experimental investigations involving QUS approaches (Barkmann et al. 2007, 2008a, 2008b, 2010; Dencks et al. 2007, 2008; Grimal et al. 2013; Grondin et al. 2010; Haiat et al. 2005; Hoffmeister et al. 2008, 2012; Karjalainen et al. 2012; Padilla et al. 2008) that obtained

encouraging results. On the other hand, to the best of our knowledge, in the literature, the only attempt to study a direct "*in vivo*" US measurement of a spine diagnostic parameter on humans is the work by Garra et al. (2009), in which spectral centroid shift was measured on vertebral bodies L3 and L4 of nine female volunteers employing a 2.5-MHz phased-array US probe.

The present study is the first to validate a different and innovative US approach to osteoporosis diagnosis applicable to the spine, integrally exploiting all features of the whole spectrum of radiofrequency (RF) signals acquired during a transabdominal echographic scan of lumbar vertebrae L1–L4 to determine the status of internal bone architecture.

The aim of this work was to introduce this new methodology and to preliminarily investigate its feasibility and accuracy, addressing two specific objectives: (i) to assess the effectiveness of a novel US-measured parameter, called the "osteoporosis score" (O.S.), in discriminating between healthy osteopenic and osteoporotic patients; (ii) to quantify the correlation between O.S. values and DXA-measured spinal BMD.

METHODS

Patients

The study was conducted at the Operative Unit of Rheumatology of "Galateo" Hospital (San Cesario di Lecce, Lecce, Italy) and included all consecutive female patients referred for a spinal DXA who fulfilled the following enrollment criteria: Caucasian ethnicity, aged 51–60 y, BMI <25 kg/m^2, absence of previous vertebral fractures, absence of significant deambulation impairment.

Enrollment criteria were established taking into account four factors:

1. The age range considered includes the majority of women referred for a spinal DXA, as younger women usually undergo osteoporosis diagnostic tests much more rarely, and older women are more frequently referred for femoral DXA because degenerative changes in the lumbar spine region may affect accuracy of the spinal scan (Chanchairujira et al. 2004; Engelke and Gluer 2006; Rand et al. 1997);
2. Low BMI is a recognized risk factor for osteoporosis and bone fractures (De Laet et al. 2005; Johansson et al. 2014; Szklarska and Lipowicz 2012), so the introduction of more effective strategies for diagnosis and prevention of osteoporosis is additionally important in the case of normal weight or underweight individuals with respect to overweight and obese persons;
3. Because the novel diagnostic approach proposed in this study requires the preliminary construction of a reference database (as detailed in subsequent subsections), we decided to gather population-based reference data, which is the approach adopted in the National Health and Nutrition Examination Survey (NHANES) III database (Looker et al. 1998) and is the preferred way to collect reference data in the field of bone densitometry (Engelke and Gluer 2006);
4. To assess the repeatability and accuracy of the proposed method, we also recruited a study population that was separate from the reference database, but fulfilled the same enrollment criteria. Therefore, we did not exclude patients with bone-related disorders (*e.g.*, secondary osteoporosis) (Engelke and Gluer 2006), and for the same reasons, we also did not exclude patients with lifestyle habits such as smoking, high or low calcium intake and sedentary life style (Botella et al. 2013; Hou et al. 2008; Pedrazzoni et al. 2003).

A total of 342 patients were recruited in 15 mo. All enrolled patients underwent two different diagnostic examinations, spinal DXA and an abdominal echographic scan of the lumbar spine, as detailed later in this section.

The study protocol was approved by the hospital ethics review board, and all patients gave informed consent.

Reference database and study population

According to widely adopted rules for establishing a reference database (Engelke and Gluer 2006; Hou et al. 2008), we grouped the enrolled patients based on their age into two 5-y intervals: 51–55 y (group A) and 56–60 y (group B). For each age interval considered, the first 100 patients were included in the reference database, and the remaining individuals represented the study population.

The size of the sample to be included in the reference database (100 patients per age group) was calculated taking into account the conclusions of the work by Hou et al. (2008), according to whom, a suitable reference database can be obtained from a population of 458 women of the same ethnicity aged 6–85 y, in which the most populated 5-y age interval includes 44 patients. We rounded off this value to 50 and multiplied it by a safety factor of 2. In this way we also respected the indications provided by Engelke and Gluer (2006), stating that sample sizes including at least 100 patients per age group can be considered fairly robust.

DXA measurements

Spinal DXA scans were performed with hip and knee both at 90° of flexion using a Discovery W scanner (Hologic, Waltham, MA, USA). BMD was measured over the lumbar spine L1–L4, and the mean value is

expressed as grams per square centimeter (g/cm^2). For each patient, Hologic software also provided the *T*-score value, defined as the number of SD from the peak BMD of young women found in the standard Hologic reference database for Caucasian women. According to the commonly used WHO definitions, patients were classified as "osteoporotic" if their *T*-score was $\leq$2.5, "osteopenic" if $<$ –2.5 but $<$ –1.0 or "healthy" if $\geq$ –1.0.

Dual X-ray absorptiometry equipment underwent daily quality control and regular maintenance for the entire study period.

Ultrasound acquisitions

Abdominal US scans of the lumbar spine were performed employing an echographic device (Echo Blaster 128, Telemed Medical Systems, Milan, Italy), equipped with a convex transducer (C3.5/60/128 Z, Telemed Medical Systems) operating at 3.5 MHz. The echographic device was provided in a research configuration that in addition to the acquisition of conventional images, also allowed the extraction of the unprocessed analogue RF signals. These signals were passed to a custom-developed signal pre-processing chain that performed the following steps: 1-kHz high-pass filtering to cancel low-frequency noise, 18-dB amplification to improve signal dynamics and analogue-to-digital conversion (40 MS/s, 16 bits).

Each patient underwent a sagittal scan of the lumbar spine, with the probe being moved back and forth from the xiphoid process. The scan lasted about 1 min and generated 100 frames of RF data (frame rate $\sim$1.5 fps) that were acquired and stored in a PC hard disk for subsequent off-line analysis. Transducer focus was set at 5 cm, and the probe was coupled with the abdomen to keep vertebral interfaces in the US focal region (*i.e.*, thickness of soft tissue between skin and vertebrae was approximately constant). Other echograph parameters were the same for all the acquisitions: power = 45%, mechanical index (MI) = 0.4, scan depth = 12 cm, gain = 0 dB, linear time gain compensation.

In the typical situation, vertebral interfaces were actually located in the focal region, being 1–2 mm below the focus reference line on the screen: L3 and L4 required almost no pressure on the probe to be placed at such a distance, whereas slight abdominal compression was necessary to reach the same placement for L1 and L2. During data acquisition, the operator oriented the US transducer to place the interface of the insonified vertebra in the horizontal position (aligned with the focus reference line).

Ultrasound scans of patients to be included in the reference database (n = 200) were performed by a sonographer experienced in abdominal investigations to ensure that the reference database included only high-quality acquisitions. Every patient in this group underwent a single US investigation.

Patient acquisitions to be included in the study population were then performed with the double aim of assessing the repeatability of the proposed method and evaluating the diagnostic accuracy on data sets acquired by inexperienced operators. Therefore, for each 5-y age interval considered, patients included in the study population were further subdivided in two groups: The first 30 patients enrolled underwent two consecutive US investigations performed by an experienced operator, with patient repositioning between the scans, and the corresponding data were used for repeatability measurements; on the other hand, subsequent patients underwent a single US examination alternately performed by one of two inexperienced operators who had previously received only a 3-h specific training session, and these data were used for accuracy measurements.

In particular, regarding repeatability assessments, patients in group A underwent two consecutive US examinations both performed by the same operator and the corresponding data were used to quantify the intra-operator repeatability (*i.e.*, technique precision), whereas patients in group B underwent two consecutive US investigations performed by two different operators and the corresponding data were used to evaluate t inter-operator repeatability.

The described patient distribution is summarized in Table 1, which also includes anthropometric details of single evaluated groups.

US data analysis

US data were analyzed through a new fully automatic algorithm that performed a series of spectral and statistical analyses, involving both the echographic images and the corresponding unfiltered RF signals, and for each patient the aforementioned O.S. value was provided as final output (details on its calculation are provided later in the text).

The implemented algorithm actually performs diagnostic parameter calculations on RF signal segments corresponding to specific regions of interest (ROIs) internal to the automatically identified vertebrae. These calculations are aimed at measuring the percentage of segments whose signal spectral features correlate better with those of an osteoporotic bone model than with those of a healthy one. In this process, the algorithm compares RF spectra calculated for the considered patient with reference model spectra of healthy and osteoporotic vertebrae derived from previous US acquisitions on DXA-classified patients. Details of the algorithm's working principles and its employment in the present study are provided in the following subsections.

Table 1. Distribution of enrolled patients among reference database and study population for each age range considered and anthropometric details for each group evaluated

		Study population		
Age range (y)	Reference database (n)	Repeatability measurement (n)	Accuracy measurement (n)	Total
51–55 (group A)	100	30 (intra-operator)	45	175
Age (y)	52.9 ± 1.4*	53.0 ± 1.5	52.5 ± 1.4	52.8 ± 1.4
Body mass index (kg/m^2)	22.5 ± 1.6	22.6 ± 1.9	22.1 ± 1.8	22.4 ± 1.7
Height (cm)	161.3 ± 5.8	160.6 ± 5.5	161.9 ± 5.9	161.3 ± 5.8
Weight (kg)	58.5 ± 6.0	58.2 ± 5.0	58.0 ± 6.3	58.3 ± 5.9
56–60 (Group B)	100	30 (inter-operator)	37	167
Age (y)	58.0 ± 1.5	58.2 ± 1.4	57.8 ± 1.4	58.0 ± 1.4
Body mass index (kg/m^2)	22.7 ± 1.8	22.4 ± 1.6	22.6 ± 2.0	22.6 ± 1.8
Height (cm)	161.3 ± 6.3	162.3 ± 6.7	161.4 ± 6.8	161.5 ± 6.5
Weight (kg)	59.0 ± 6.0	58.9 ± 6.0	58.8 ± 7.0	58.9 ± 6.2
Total	200	60	82	342

* Mean ± standard deviation.

Theoretical background. The main hypothesis of the adopted approach is that RF spectra of US signals backscattered from lumbar vertebrae, and trans-abdominally acquired, contain useful information about the corresponding bone status, as also previously reported (Garra et al., 2009).

Actually, it has been widely reported that frequency spectra of US backscatter signals are quantitatively related to the characteristic acoustic properties of the scattering medium, such as backscatter coefficient $\eta(f)$, speed of sound c and attenuation coefficient $\alpha(f)$, where f = frequency (Hoffmeister 2011; Hoffmeister et al. 2012; O'Donnell and Miller 1981; Sigelmann and Reid 1973). In particular, by use of the notation adopted by Hoffmeister et al. (2012), the power spectrum of the US signal backscattered from a cancellous bone specimen immersed in water, in the single-scattering assumption, can be expressed as

$$P(f) = \frac{4P_0(f)\cdot E(f)\cdot V(f)\cdot \eta(f)\cdot T^2}{r^2\cdot A(f)} \qquad (1)$$

where $P_0(f)$ is the power spectrum of the transmitted pulse, $E(f)$ is the two-way electromechanical power conversion efficiency of the measurement system, $V(f)$ is the gated scattering volume of the specimen (*i.e.*, the bone volume that produced the backscattered signal portion gated through the analysis window of duration τ_w as subsequently defined), T is the intensity transmission coefficient at the water–specimen interface, r is the distance from the transducer to the center of the scattering volume and $A(f)$ is an attenuation term given by

$$A(f) = e^{4\alpha(f)x}\left[\frac{2\alpha(f)c\tau_w\cdot e^{\alpha(f)c\tau_w}}{e^{\alpha(f)c\tau_w} - e^{-\alpha(f)c\tau_w}}\right] \qquad (2)$$

where x is the intervening bone thickness between the water–specimen interface and the start of the analysis gate, and τ_w represents the analysis window duration: $A(f)$, in fact, includes the ultrasonic pulse attenuation up to the start of the analysis gate and within the gated region (Hoffmeister et al. 2012).

In our study we aimed at improving the robustness of *in vivo* transabdominal US assessment of lumbar spine by combining several assumptions involving equations (1) and (2) with, first, a fully automatic selective identification (see following paragraphs) of vertebral interfaces and underlying trabecular regions to be analyzed (to make the method more objective and operator independent) and, second, a novel statistical approach in data processing (to increase the diagnostic discrimination power by reducing the noise impact).

In more detail, we employed the same echographic device with the same settings for all reported acquisitions; therefore, in equation (1), $P_0(f)$ and $E(f)$ were assumed to be constant functions of f; the analysis window had a fixed length, implying that τ_w in equation (2) had a constant value; the distance between the US transducer and the vertebral interfaces was kept to a constant value; and the trabecular analysis region was always selected with the same method, so we assumed that $V(f)$, r and x were also constant. Within the aforementioned first-approximation assumptions, and temporarily neglecting the effects of soft tissues between the transducer and the vertebral interfaces, we can say that the backscattered US power spectrum $p(f)$ depends only on $\eta(f)$, T, $\alpha(f)$ and c, where T is the only term that even within the above-mentioned assumptions, is still dependent on factors different from vertebra properties (because T is related to the acoustic impedance both of the vertebra and of the encountered tissues).

To account for the effects of the soft tissue between transducer and spine, our analysis included the entire lumbar spine tract L1–L4, and a particular series of averaging and normalization operations was implemented, as

detailed in the next subsections, to account for local variations. Additionally, the constant depth value of analyzed vertebrae was chosen in relation to the BMI range of the patients studied to require minimal or no pressure to be reached and, therefore, minimizing the effect on sound speed variations in soft tissue. Furthermore, the interface of each analyzed vertebra had to be placed horizontally in the image (to minimize the inclination effects on backscattered signals).

Our approach was finally completed by two characteristic features: (i) a custom-implemented algorithm was employed for fully automatic selective identification of vertebrae and the corresponding analysis region (described in a dedicated subsection and detailed in the Appendix); (ii) each calculated trabecular spectrum underwent an overall shape comparison with two age-matched reference models that had been calculated from a database of real osteoporotic and healthy patient data, because our aim was to improve the correlations between US backscatter parameters and DXA-measured spinal BMD taking into account the whole spectrum of characteristics rather than single specific features (*e.g.*, centroid shift) and comparing them with patient-based models instead of reference values measured on phantoms. In fact, preliminary tests were performed using alternatively specific features of the spectra (*e.g.,* specific bandwidth portions, spectrum slopes, spectrum center frequencies), but the correlation values were always inferior to that provided with the above-described total spectrum approach. The number of patients to be included in each model was specifically optimized for the adopted processing chain, to maximize the discrimination power between osteoporotic, osteopenic and healthy patients (further details are provided in the following subsections).

Brief overview of the adopted analysis methodology. In short, our proposed approach is based on a highly selective automatic identification of vertebrae and ROIs combined with statistical shape comparisons between selected frequency spectra of RF signals backscattered from patient vertebrae and reference spectral models, which had been previously derived from age-matched "osteoporotic" and "healthy" patients through an innovative method detailed later.

Data analysis performed on each single patient can be summarized in five main steps:

1. Automatic and selective identification of vertebrae within the sequence of acquired echographic images;
2. For each identified vertebra, automatic identification of a specific portion of RF signal for each scan line crossing the vertebra surface;
3. Classification of the vertebral region corresponding to each considered RF signal portion as "osteoporotic" or "healthy" based on the degree of correlation of its frequency spectrum with the pair of age-matched spectral reference models that had been previously derived (an "osteoporotic" one and a "healthy" one);
4. For each vertebra considered, calculation of the previously mentioned O.S. value (*i.e.*, the novel diagnostic parameter introduced) defined as the percentage of the analyzed vertebra regions that were classified as "osteoporotic" in the previous step (more details on O.S. calculation are provided later under Automatic Identification of Vertebrae and Calculation of Osteoporosis Score; see in particular equations 9 and 10)
5. Calculation of the final O.S. value for the patient considered as the average of the values calculated for single vertebrae.

Calculation of reference models. For both 5-y age intervals considered (51–55 y and 56–60 y), US data sets from patients included in the reference database were used to calculate the corresponding model spectra through the following procedure (also schematically illustrated in Fig. 1).

The US data of all the patients classified as "osteoporotic" on the basis of DXA diagnosis were processed through six steps to obtain the related "osteoporotic model":

1. For the *k*th osteoporotic patient considered, four echographic images of the lumbar vertebrae L1–L4 were manually selected, with each vertebra image including E_i echographic lines ($i = 1, \ldots, 4$) (at this step an experienced operator chose, for each lumbar vertebra from L1 to L4, the best available image frame, in terms of both signal-to-noise ratio and actual satisfaction of interface placement requirements, manually identifying the E_i echographic lines crossing the interface);
2. For the *j*th echographic line crossing the *i*th selected vertebra ($j = 1, \ldots, E_i$, $i = 1, \ldots, 4$), a segment of the corresponding RF signal was selected through a 200-point Hamming window located immediately after the echo from the vertebral surface and zero-padded to 4096 points (the Hamming window started when the amplitude of the RF signal envelope reached 15% of its peak value corresponding to the bone interface);
3. The fast Fourier transform (FFT) power spectrum $P_{\mathrm{Exp}_{ij}}(f)$ was calculated (in dB) for the obtained *j*th signal from the *i*th selected vertebra and compensated to take into account the frequency sensitivity of the US transducer employed (*i.e.*, the probe transfer function $P_{\mathrm{TF}}(f)$) according to the equation

$$P_{\mathrm{Comp}_{ij}}(f) = P_{\mathrm{Exp}_{ij}}(f) - P_{\mathrm{TF}}(f) \tag{3}$$

where $P_{\mathrm{Comp}_{ij}}(f)$ is the compensated spectrum. The probe transfer function $P_{\mathrm{TF}}(f)$ was determined in a water tank

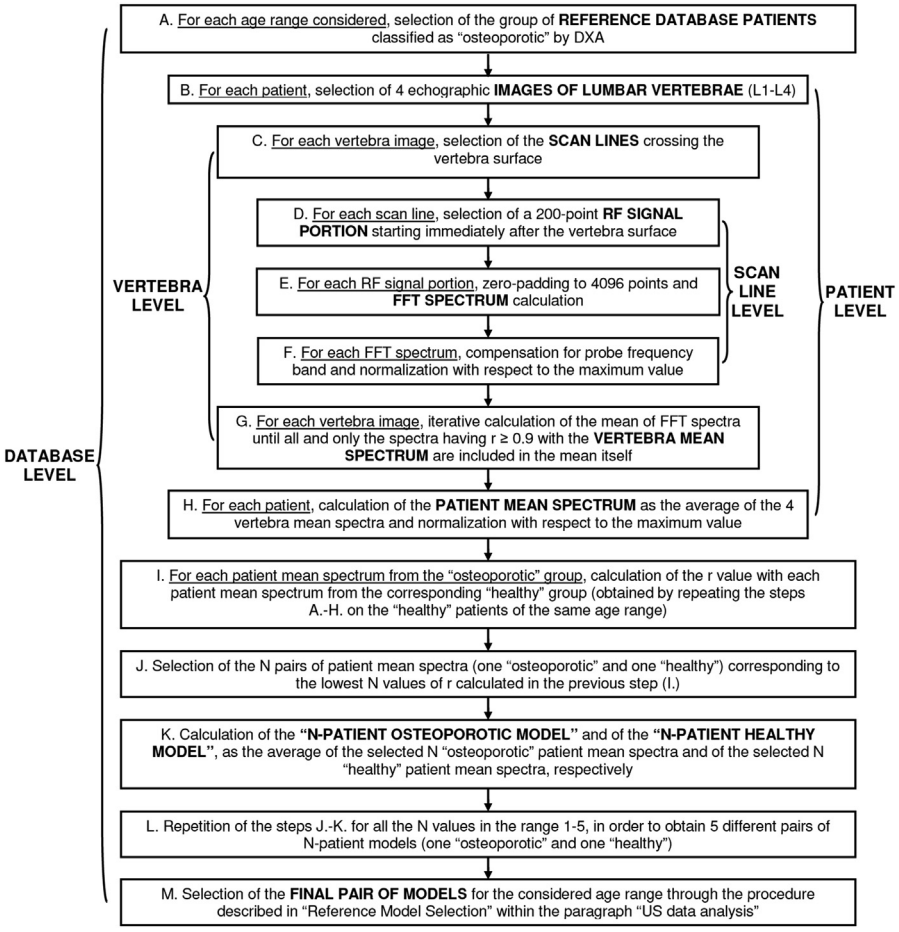

Fig. 1. Schematic illustration of the data processing steps implemented to calculate the pairs of reference model spectra for each 5-y age range. DXA = dual X-ray absorptiometry; FFT = fast Fourier transform; RF = radiofrequency.

by transmitting a broadband US pulse toward a steel plate placed at the focal distance and calculating the FFT spectrum of the reflected signal;

4. The obtained spectrum $P_{\mathrm{Comp}_{ij}}(f)$, referred to the *j*th echographic line of the *i*th vertebra, was normalized with respect to its maximum value:

$$P_{\mathrm{Norm}_{ij}}(f) = P_{\mathrm{Comp}_{ij}}(f) - \max_f \left[P_{\mathrm{Comp}_{ij}}(f) \right] \qquad (4)$$

5. The spectrum representing the *i*th vertebra (called "vertebra mean spectrum" $\overline{P}_i(f)$) was obtained through the following iterative procedure:
   a. Initial calculation of the vertebra mean spectrum, estimated from the spectrum averaged over all the echographic lines via the formula

$$\overline{P}_{i_0}(f) = \frac{\sum_{j=1}^{E_i} P_{\mathrm{Norm}_{ij}}(f)}{E_i} \qquad (5)$$

   b. Calculation of the Pearson correlation coefficient r_{ij_0} between $\overline{P}_{i_0}(f)$ and each $P_{\mathrm{Norm}_{ij}}(f)$ (r_{ij_0} is calculated in the range 1–5 MHz)
   c. Identification, among the spectra $P_{\mathrm{Norm}_{ij}}(f)$, of the E_{i_1} spectra having $r_{ij_0} \geq 0.90$ ($E_{i_1} \leq \mathrm{E_i}$)
   d. Update of the vertebra mean spectrum calculation to include only the E_{i_1} spectra identified in the previous step:

$$\overline{P}_{i_1}(f) = \frac{\sum_{j=1}^{E_{i_1}} P_{\mathrm{Norm}_{ij}}(f)}{E_{i_1}} \qquad (6)$$

   e. Iteration of steps b to d until achievement of a vertebra mean spectrum $\overline{P}_i(f)$ including all and only those spectra $P_{\mathrm{Norm}_{ij}}(f)$ having $r_{ij} \geq 0.90$ with the average spectrum $\overline{P}_i(f)$

6. The four spectra $\overline{P}_i(f)$ obtained (one for each vertebra, $i = 1, \ldots, 4$) were averaged to obtain a spectrum that was normalized with respect to its maximum value

and assumed as representative of the kth osteoporotic patient considered ("osteoporotic patient mean spectrum" $\overline{PS}_{\mathrm{Ost}_k}(f)$):

$$\overline{PS}_{\mathrm{Ost}_k}(f) = \frac{\sum_{i=1}^{4} \overline{P}_i(f)}{4} - \max_f \left(\frac{\sum_{i=1}^{4} \overline{P}_i(f)}{4} \right) \quad (7)$$

The US data from the "healthy" patients were processed in an analogous manner, until for each zth healthy patient, a representative "healthy patient mean spectrum" $\overline{PS}_{\mathrm{Heal}_z}(f)$ was obtained. In this way, each "osteoporotic" or "healthy" patient was represented by a single spectrum. For each 5-y age interval, we then considered one by one the "osteoporotic patient mean spectra" $\overline{PS}_{\mathrm{Ost}_k}(f)$ and calculated the Pearson correlation coefficient r_{kz} between the considered "osteoporotic" spectrum $\overline{PS}_{\mathrm{Ost}_k}(f)$ and each "healthy patient mean spectrum" $\overline{PS}_{\mathrm{Heal}_z}(f)$ obtained from the "healthy" group of patients belonging to the same age interval. Among all the calculated r_{kz} values, the lowest N values were selected, which identified the N pairs of spectra showing the minimum mutual correlation. The N corresponding "osteoporotic" spectra were averaged to obtain a spectrum that was normalized with respect to its maximum value and labeled as the "N-patient osteoporotic model" $\overline{\mathrm{MS}}_{\mathrm{Ost}_{N_y}}(f)$ for the considered age interval (subscript y identifies the age interval: 51–55 y or 56–60 y):

$$\overline{\mathrm{MS}}_{\mathrm{Ost}_{N_y}}(f) = \frac{\sum_{k=1}^{N} \overline{PS}_{\mathrm{Ost}_k}(f)}{N} - \max_f \left(\frac{\sum_{k=1}^{N} \overline{PS}_{\mathrm{Ost}_k}(f)}{N} \right) \quad (8)$$

An "N-patient healthy model" $\overline{\mathrm{MS}}_{\mathrm{Heal}_{N_y}}(f)$ was analogously obtained for each age interval by considering the N "healthy" spectra of the N couples of spectra showing the minimum mutual correlation. N was varied in the range 1–5, obtaining five different pairs of models for each age interval. Each pair was composed of an "N-patient healthy model" $\overline{\mathrm{MS}}_{\mathrm{Heal}_{N_y}}(f)$ based on N healthy patients and an "N-patient osteoporotic model" $\overline{\mathrm{MS}}_{\mathrm{Ost}_{N_y}}(f)$ based on N osteoporotic patients ($N = 1, \ldots, 5$).

Reference model selection. For each age interval, each of the five pairs of models obtained ($\overline{\mathrm{MS}}_{\mathrm{Ost}_{N_y}}(f)$, $\overline{\mathrm{MS}}_{\mathrm{Heal}_{N_y}}(f)$) was employed in the procedure described in the next subsection to calculate the O.S. value for all reference database patients belonging to the age interval considered and not used for model construction. The "optimal" pair of models ($\overline{\mathrm{MS}}_{\mathrm{Ost}_{\overline{N}_y}}(f)$, $\overline{\mathrm{MS}}_{\mathrm{Heal}_{\overline{N}_y}}(f)$) was finally selected based on the best discrimination between osteoporotic, osteopenic and healthy patients derived from the O.S. values obtained. In this way, a different couple of model spectra were associated with each age interval considered, together with specific O.S. diagnostic thresholds that were determined to maximize the discrimination power among osteoporotic, osteopenic and healthy patients.

The next subsection describes the typical operations performed by the algorithm on the US data acquired on a generic patient to automatically identify vertebrae and calculate the O.S. value exploiting the selected reference models.

Automatic identification of vertebrae and calculation of osteoporosis score. The first operation carried out by the algorithm is the automatic identification of vertebral interfaces within the sequence of echographic images acquired on the kth patient considered. This is achieved by performing steps 1–9 on each acquired frame according to the related indicated criteria (the corresponding block diagram is given in Fig. 2; Fig. 3 comprises two sample echographic images, one of suitable quality [Fig. 3a] and "noisy" one [Fig. 3b]; Fig. 4

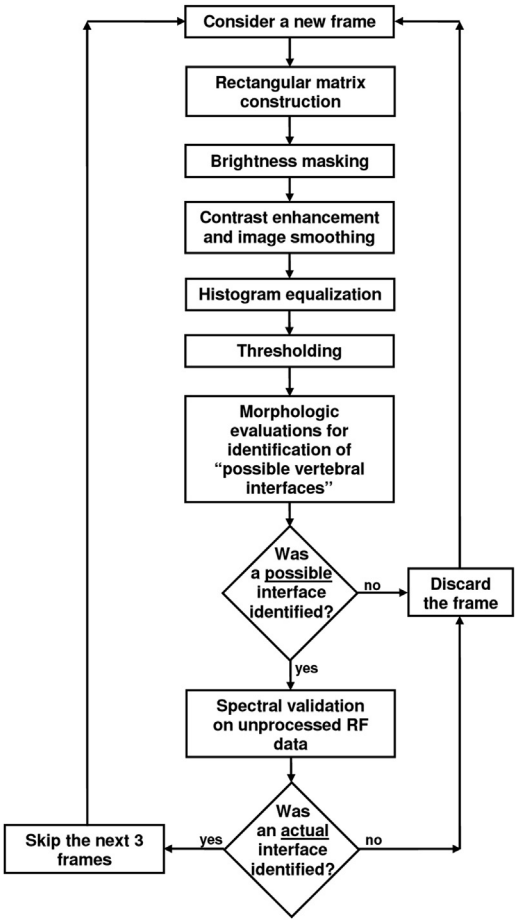

Fig. 2. Schematic illustration of the automatic identification of vertebral interfaces within the sequence of ultrasound data frames acquired on each patient. RF = radiofrequency.

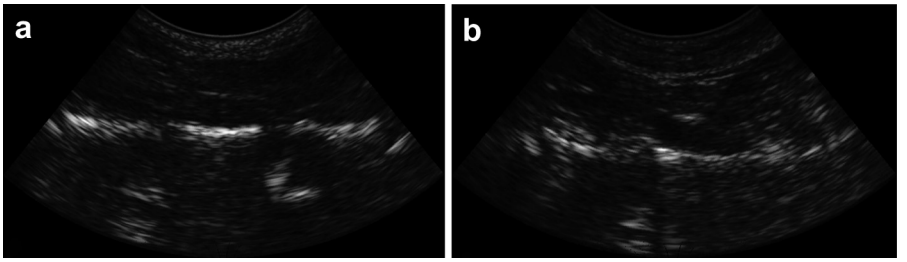

Fig. 3. Sample echographic images: (a) Example of a frame containing a vertebral interface that will be automatically identified by the algorithm. (b) Example of a "noisy" frame that will be automatically discarded by the algorithm because there are no vertebral interfaces suitable for diagnostic analysis.

illustrates how steps 1–9 are applied to the image frames reported in Fig. 3 to automatically identify the vertebra in Fig. 3a and to discard the "noisy" frame in Fig. 3b; the single image processing steps are simply listed here together with the references to different parts of Fig. 4 to provide a qualitative visualization of the results of each step; quantitative details on the corresponding operations are reported in the Appendix):

1. Rearrangement of image data in a rectangular matrix (Fig. 4a, b);
2. Brightness masking (Fig. 4c, d);
3. Contrast enhancement and image smoothing (Fig. 4e, f);
4. Histogram equalization (Fig. 4g, h);
5. Thresholding (Fig. 4i, j);
6. Morphologic evaluations (Fig. 4k, l);
7. Spectral validation (performed only if there are clusters of white pixels retained after the previous step and consisting in the fact that spurious interfaces are filtered out if no correlation is found with either reference model; see Appendix for details);
8. Skip of the subsequent three frames (only if an "actual vertebral interface" has been identified in the previous step; see Appendix for details);
9. New analysis started on the next frame.

Once the described process has been iterated until all the frames belonging to the US data set of the kth patient have been analyzed, a number n_k of vertebral interfaces has been identified; if $n_k < 4$, the data set is labeled as "noisy" and the O.S. value is not calculated. Otherwise, the algorithm proceeds to the following diagnostic calculations on each RF spectrum of the identified ROIs, selected as in the case of model construction (200-point Hamming-windowed signal portions starting immediately after the echo from the vertebral surface, when the amplitude of RF signal envelope reached 15% of its peak value).

The jth RF spectrum of the ith ROI ($S_{\mathrm{Norm}_{ij}}(f)$) is classified as "osteoporotic" if the value of its Pearson correlation coefficient $r_{\mathrm{Ost}_{ij}}$ with the appropriate osteoporotic model spectrum (*i.e.*, the age-matched $\overline{\mathrm{MS}}_{\mathrm{Ost}_{\overline{N}_y}}(f)$) is higher than the corresponding $r_{\mathrm{Heal}_{ij}}$ value with the related healthy model spectrum (*i.e.*, the age-matched $\overline{\mathrm{MS}}_{\mathrm{Heal}_{\overline{N}_y}}(f)$); otherwise, it is classified as "healthy." The number of spectra classified as "osteoporotic" for the ith vertebra was labeled $E_{i_{\mathrm{ost}}}$ (with $E_{i_{\mathrm{ost}}} \leq E_i$, where E_i is always the number of echographic lines crossing the vertebral surface, coinciding with the total number of RF spectra belonging to the ROI). Then, the percentage of analyzed spectra classified as "osteoporotic" represents the O.S. of the considered vertebra:

$$O.S._{V_i} = \frac{E_{i_{\mathrm{ost}}}}{E_i} \cdot 100 \tag{9}$$

The same evaluations are repeated for each identified vertebra, and the final O.S. related to the kth patient analyzed is obtained as the average of the O.S. values calculated for single vertebrae:

$$\mathrm{O.S.}_k = \frac{\sum_{i=1}^{n_k} \mathrm{O.S.}_{V_i}}{n_k} \tag{10}$$

where n_k represents the number of vertebrae identified in the US data set of the kth patient.

In the first part of the present study, the illustrated procedure was employed to analyze the data sets of the patients included in the reference database, to choose the best couple of models ($\overline{\mathrm{MS}}_{\mathrm{Heal}_{\overline{N}_y}}(f)$, $\overline{\mathrm{MS}}_{\mathrm{Ost}_{\overline{N}_y}}(f)$) for each age interval considered. Once the model selection step was completed, the same procedure was used to assess the actual effectiveness of the selected models through independent tests on study population patient data sets. These tests first evaluated method repeatability (*i.e.*, intra- and inter-operator variability) and then assessed its diagnostic discrimination power with respect to DXA, assumed as a gold standard reference, as detailed next.

The algorithm employed in this study was implemented in MATLAB (The MathWorks, Natick, MA, USA) and run on a personal notebook computer equipped with an Intel i7 processor at 2.3 GHz and 8 GB of RAM. The US data analysis of a single patient, including automatic identification of vertebrae and calculation of O.S.

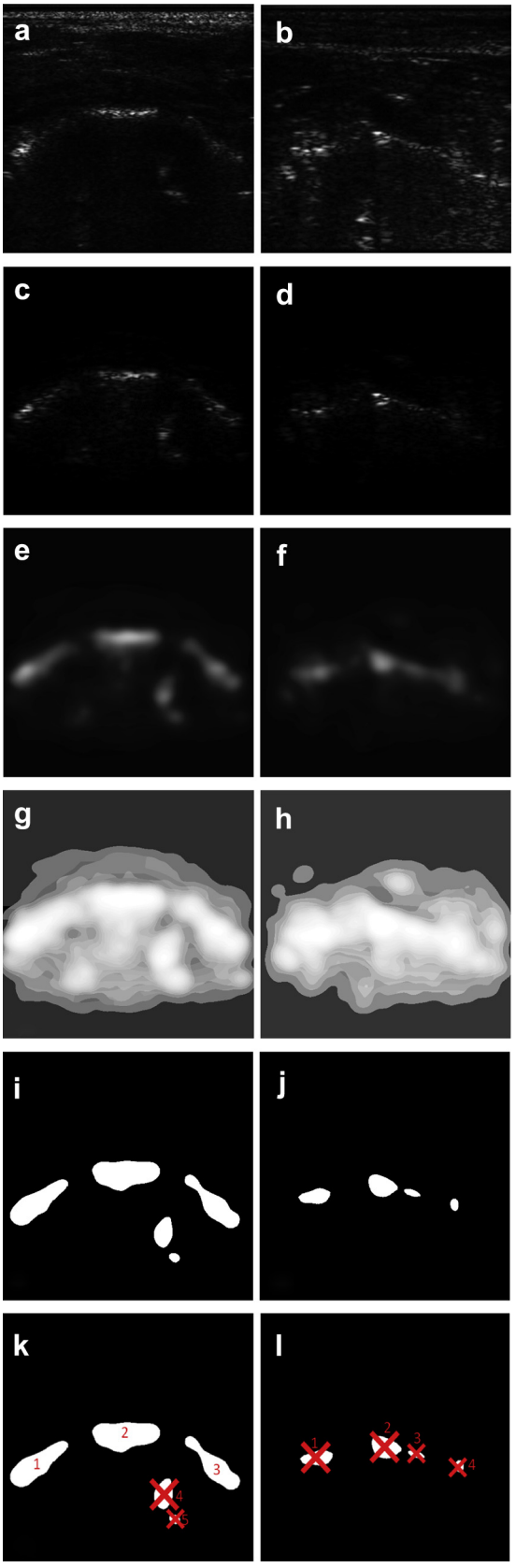

Fig. 4. Application of the processing steps for automatic vertebral interface identification to the frames in Figure 3. Images on the left refer to the frame in Figure 3a, and images on the right, to the frame in Figure 3b. (a, b) Rearrangement of image data in rectangular matrices. (c, d) After brightness masking. (e, f) After contrast enhancement and image smoothing. (g, h) After histogram equalization. (i, j) After thresholding. (k) Morphologic evaluations: Clusters 4 and 5 are excluded because they are outside the expected size range, whereas clusters 1–3 are retained, and after the morphologic evaluations described in the Appendix, cluster 2 is labeled as a "possible vertebral interface." (l) Morphologic evaluations: All the clusters are excluded because they are outside the expected size range, and the frame is discarded.

value employing a given pair of models, was completed within 2 min.

Statistical analysis

Reference model selection. To choose the pair of models providing the best diagnostic power, for each age range considered, patients included in the reference database were further subdivided into three groups on the basis of DXA diagnosis (osteoporotic, osteopenic and healthy), and the effectiveness of each tested pair of models was evaluated through the methods described next.

For every pair of models tested ($\overline{\mathrm{MS}}_{\mathrm{Heal}_N}(f)$, $\overline{\mathrm{MS}}_{\mathrm{Ost}_N}(f)$, $N = 1, \ldots, 5$), after the O.S. value had been calculated for each patient, specific O.S. diagnostic thresholds were automatically established as the values providing the best discrimination between patients of different DXA-based groups and a new patient subdivision was operated according to this O.S.-based diagnostic criterion. Level of agreement between the two methods of identifying osteoporotic, osteopenic and healthy patients was assessed through the calculation of accuracy (*i.e.*, correct diagnoses/patients analyzed) and Cohen's κ (Cohen 1960).

For each age interval considered, the pair of models providing the best performance according to the two aforementioned parameters was selected, together with the corresponding O.S. diagnostic thresholds, to be employed for the subsequent tests on the study population.

Intra-operator variability. Intra-operator variability was assessed in terms of "short-term precision" as defined by Engelke and Gluer (2006) using the data acquired on the first 30 patients of group A included in the study population.

For each individual considered, the SD of the repeated US measurements was calculated and method precision was then expressed as the root mean square (RMS) average of the SD (RMS-SD) and as the root-mean-square coefficient of variation (RMS-CV). Least significant change (LSC) for a 95% confidence level

was also calculated as recommended by the ISCD. These calculations were performed via the ISCD precision calculator (available at http://www.iscd.org/resources/calculators/). To facilitate the interpretation of numerical data and comparison with previously published articles, all the results related to precision and repeatability of the proposed methods were expressed in terms of BMD (g/cm^2) by employing a linear regression approach to estimate spinal BMD from O.S. values, as described later in the text.

Inter-operator variability. Inter-operator variability was assessed on the data acquired on the first 30 patients of group B included in the study population. Quantifications were carried out similarly to those for intra-operator variability: For each individual considered, the SD of the repeated US measurements was calculated, and inter-operator variability was then expressed as RMS-SD and as RMS-CV. LSC for a 95% confidence level was also calculated. The inter-observer correlation was also assessed using Cohen's κ (Cohen 1960).

Accuracy of the novel US method compared with DXA. Diagnostic accuracy of the proposed US method with respect to DXA measurements, assumed as a gold standard reference, was separately evaluated on study population patients not used for repeatability assessments. In this case, acquired US data were analyzed employing only the selected age-matched couple of reference models and the corresponding O.S. diagnostic thresholds, both selected as described under Reference Model Selection.

To assess the effectiveness of O.S. values in discriminating between osteoporotic, osteopenic and healthy patients, in each age range considered, patients were further subdivided into these three corresponding groups on the basis of DXA diagnosis, and the mean and SD of O.S. values were calculated for each patient group. Statistical significance of the differences between the calculated mean values was evaluated using an unpaired two-tailed Student *t*-test.

Level of agreement between the two methods of classifying patients was assessed by calculation of accuracy and Cohen's κ, similar to what had been done for the reference model. To quantify the correlation between O.S. values and DXA-measured BMD, Pearson's correlation coefficient (r), the coefficient of determination (r^2) and residual errors (RMSE) were also calculated in each age range, employing a linear regression approach to estimate spinal BMD from O.S. values. Furthermore, the agreement between DXA-measured BMDs and US-derived values was assessed, as recommended by Altman and Bland (1983), by calculating the paired difference for each measurement and by estimating the bias (mean difference) and 95% limits of agreement (2 SD around the mean difference) relative to the average measurement of both methods.

RESULTS

Reference model selection

For each age range considered, data sets of patients included in the reference database were tested with five different pairs of models ($\overline{MS}_{Heal_{N_y}}(f)$, $\overline{MS}_{Ost_{N_y}}(f)$, $N = 1, \ldots, 5$), determining each time the specific O.S. diagnostic thresholds that provided the best discrimination between osteoporotic, osteopenic and healthy patients as classified by DXA, assumed as a gold standard reference. A comparative diagnostic performance evaluation of different tested models was then performed through the calculation of accuracy and Cohen's κ.

Accuracy trends as a function of the number of patients included in the model are illustrated in Figure 5. For both age intervals, maximum accuracy was reached with the employment of models based on three patients (96.8% in group A, 94.7% in group B). This behavior can be explained by the fact that a model based on a single patient is likely to be highly influenced by the specific features of that patient, being not very suitable for a population diagnosis, whereas the progressive inclusion of more patients in the model increases the number of specific features taken into account, giving a greater weight to the features that are common among the included patients and reducing the importance of those related only to one patient. On the other hand, the increasing number of patients included in the models also causes an increased degree of similarity between the "osteoporotic" model and the corresponding "healthy" one, with a consequent reduction in discrimination power. In our reference database, for both age intervals, the best

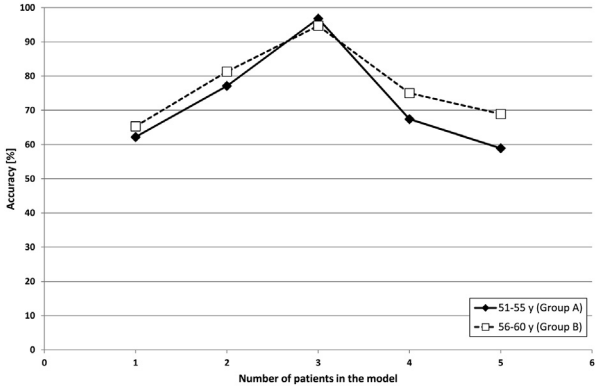

Fig. 5. Accuracy as a function of the number of patients included in each model. (Accuracy calculations were conducted on the age-matched patients included in the reference database and not used for model construction.)

compromise was the adoption of models based on three patients each.

This was further confirmed by the corresponding Cohen κ values, reported in Table 2. Actually, the highest values were again found in correspondence of the three-patient models, and their interpretation was unmistakable; according to the criteria reported by Landis and Koch (1977), in fact, a κ value >0.80 indicates "almost perfect agreement" between DXA- and US-based patient classifications, and based on the criteria reported by Fleiss (1981), a κ value >0.75 signifies "excellent reproducibility" of the obtained results. In fact, as expected, Cohen's κ trends reported in Table 2 resembled the accuracy trends illustrated in Figure 5, with a more evident peak in correspondence with $N = 3$: This is clearly a result of the differences between the mathematical formulation of accuracy (whose result simply represents a portion of the analyzed cases and is therefore limited to the range between 0 and 1) and Cohen's κ formula (which can also provide negative values for methods performing worse than what could be expected on the basis of pure chance). From a practical point of view, κ values tend to be high for methods with very good accuracy levels, whereas they decrease more rapidly than accuracy when method performance gets worse.

Therefore, for both age intervals, we selected the pair of models based on three patients, together with the corresponding O.S. diagnostic thresholds, to be employed for the subsequent tests on study population data sets. The selected couple of models are illustrated in Figure 6. For both age intervals, the most evident difference between the healthy and osteoporotic models is that in the former, the higher frequencies are more attenuated, because it is obtained from the spectra of signals backscattered from healthy bones, which are less porous and more attenuating than those backscattered from osteoporotic b. Therefore, as an overall effect, healthier bones act as a kind of low-pass filter on the backscattered US signal,

Table 2. Cohen's κ values expressing the level of agreement between ultrasound and dual X-ray absorptiometry diagnoses of reference database patients as a function of the number of patients included in each model

Number of patients in model (N)	Age range (y)	
	51–55 (group A)	56–60 (group B)
1	0.416†	0.434†
2	0.647†	0.693†
3	0.945†	0.909†
4	0.455†	0.602†
5	0.315*	0.511†

* $p < 0.001$.
† $p < 0001$.

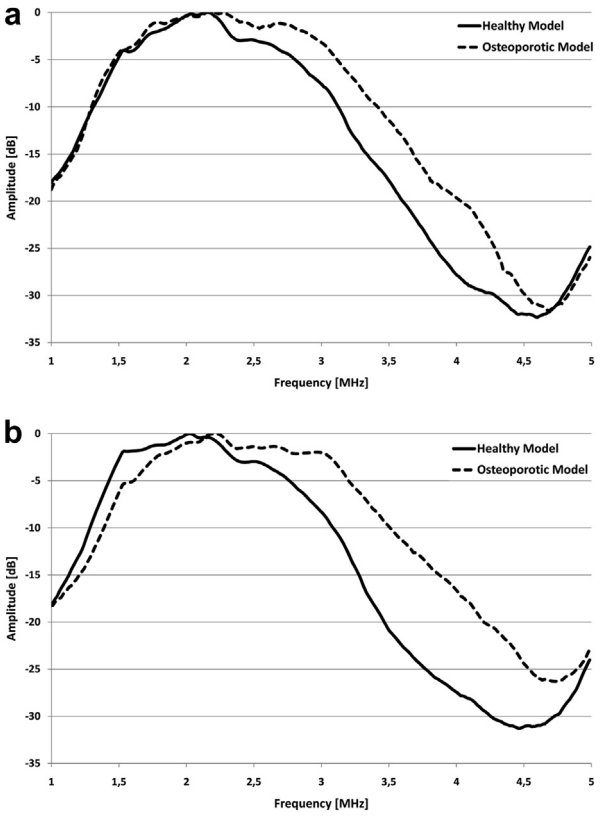

Fig. 6. Finally selected reference model spectra: $\overline{\mathrm{MS}}_{\mathrm{Heal}_{\overline{N_y}}}(f)$, $\overline{\mathrm{MS}}_{\mathrm{Ost}_{\overline{N_y}}}(f)$. (a) Healthy model and osteoporotic model for patients aged 51–55 y (group A). (b) Healthy model and osteoporotic model for patients aged 56–60 y (group B). (Each model included data from three patients, as described in the text.)

and this effect seemed to be more evident for the older patients (group B, Fig. 6b). In addition, it is interesting to note that the described behavior implies that there are fairly wide frequency ranges in which backscatter provided by the denser healthy bones is markedly inferior to the corresponding response from osteoporotic bones; this is actually not surprising, because attenuation effects may play an important role as bone density increases, overtaking the simultaneous increments in backscatter coefficient (Hoffmeister 2011). However, the proposed approach is further intrinsically characterized by the fact that all the model shape features are taken into account in the calculation of O.S. values; additional possible interpretations of the trends reported in Figure 6 are provided in the Discussion section. The O.S. thresholds related to the models illustrated in Figure 6, which provided the reported diagnostic performances, are as follows: patients aged 51–55 y (group A) were classified as "osteoporotic" if their O.S. was ≥57, "osteopenic" if >46 but <57 or "healthy" if ≤46; patients aged 56–60 y (group B) were classified as "osteoporotic" if their O.S. was ≥56, "osteopenic" if >44 but <56 or "healthy" if ≤44.

Table 3. Inter-operator variability of the proposed ultrasound-based method: Cross-tabulation of the diagnostic classifications of 30 patients independently investigated by two different operators

	Operator A			
Operator B	Healthy	Osteopenic	Osteoporotic	Total
Healthy	7	1	0	8
Osteopenic	1	14	0	15
Osteoporotic	0	1	6	7
Total	8	16	6	30

Precision and repeatability

Precision assessments (*i.e.*, intra-operator variability) and inter-operator repeatability tests were aimed at assessing the intrinsic precision and reproducibility of the proposed method, without any possible influence of operator experience.

Precision, expressed as RMS-SD, was 0.028 g/cm^2 (RMS-CV = 2.95%), and the corresponding LSC for a 95% confidence level was 0.078 g/cm^2 (RMS-CV = 8.17%). Analogous calculations were performed to assess inter-operator variability, producing the following results: RMS-SD = 0.033 g/cm^2 (RMS-CV = 4.00%), LSC = 0.092 g/cm^2 (RMS-CV = 11.09%).

Inter-observer correlation was also assessed through Cohen's κ, with $\kappa = 0.838$ ($p < 0.0001$) optimal accord between the operators. To further emphasize the level of agreement between the two operators, the corresponding data are cross-tabulated in Table 3.

Accuracy of the novel US method compared with DXA

Accuracy measurements were conducted on a separate group of patients included in the study population (see Table 1), whose US scans were performed by inexperienced operators, to also assess the actual user-friendliness of the proposed approach and the effectiveness of the algorithm in the automatic identification of "noisy" acquisitions.

Application of the described algorithm to the analysis of these US data sets provided the preliminary automatic identification of three "noisy" acquisitions (3.7% of study population patients initially enrolled for accuracy measurements), which were excluded from subsequent O.S. calculations and diagnostic analyses. Therefore, the great majority of US acquisitions performed by inexperienced operators (79/82, 96.3%) were of suitable quality for actual diagnostic calculations. In particular, two of the excluded acquisitions corresponded to patients from group A (4.4% of those initially enrolled for this purpose) and one to a patient from group B (2.7%). Two of the "noisy" acquisitions were carried out by the same operator and one by the other; thus, acquisitions performed by the first inexperienced operator (n = 41) resulted in suitable quality in 95.1% of cases and acquisitions performed by the second one (n = 41) resulted in suitable quality in 97.6% of cases.

Table 4 summarizes the distribution of the study population patients enrolled for accuracy measurements and whose US data sets were automatically judged to be of suitable quality by the algorithm. For each age range considered, patients are subdivided into three diagnostic categories (osteoporotic, osteopenic, healthy) according to DXA measurements and, for each patient group. Table 4 also lists the average values of DXA-measured BMDs and US-based O.S. values, together with the statistical significance of the difference between each reported mean O.S. value and those obtained for patients belonging to different DXA-based groups in the same age range.

In each age range considered, O.S. values of osteoporotic patients were significantly higher than the corresponding values of either osteopenic or healthy patients, and moreover, O.S. values of healthy patients were significantly lower than those of osteopenic patients ($p < 0.0001$ for all the aforementioned differences in O.S. values).

Therefore, each enrolled patient was again classified as osteoporotic, osteopenic or healthy by applying the appropriate O.S. thresholds that had been previously identified for each age range; patients who received the

Table 4. Results of DXA measurements and ultrasound-based osteoporosis scores for patients used in accuracy measurements*

		DXA			
Age range (y)	Number of patients analyzed	DXA diagnosis	n	Bone mineral density (g/cm^2)	Ultrasound-based osteoporosis score
51–55 (group A)	43	Osteoporosis	8	0.730 ± 0.020†	57.4 ± 1.9‡
		Osteopenia	21	0.838 ± 0.040	51.4 ± 2.8‡
		Healthy	14	1.002 ± 0.041	44.0 ± 3.7‡
56–60 (group B)	36	Osteoporosis	15	0.711 ± 0.068	57.2 ± 2.6‡
		Osteopenia	15	0.858 ± 0.046	52.7 ± 2.3‡
		Healthy	6	1.049 ± 0.075	40.5 ± 2.1‡

DXA = dual X-ray absorptiometry.
* Patients are grouped on the basis of age range and DXA diagnosis, excluding those who were automatically identified as "noisy acquisitions."
† Mean ± SD.
‡ $p < 0.0001$.

Table 5. Diagnostic effectiveness of ultrasound-based osteoporosis score: Agreement with DXA results for the patients used in accuracy measurements

Age range (y)	Number of patients analyzed	Agreement between ultrasound and DXA	
		Accuracy	Cohen's κ
51–55 (group A)	43	90.7%	0.841*
56–60 (group B)	36	91.7%	0.867*
Total	79	91.1%	0.859*

DXA = dual X-ray absorptiometry.
* $p < 0.0001$.

same classification in both DXA and US systems were identified as "correct diagnoses." Accuracy levels and corresponding values of Cohen's κ are listed in Table 5. Accuracy was always above 90%, which together with values of Cohen's κ always well above 0.80, documented very good agreement between DXA and US-based diagnoses along the entire age interval considered. Additionally, no patient was diagnosed as "healthy" by one technique and as "osteoporotic" by the other.

Furthermore, employing a linear regression approach, we used O.S. values to estimate spinal BMD. In each age range, the BMD estimates obtained exhibited appreciable and statistically significant correlation with the DXA-measured values: $r = 0.86$ in group A ($p < 0.001$) and $r = 0.82$ in group B ($p < 0.001$). The corresponding high values of the coefficient of determination ($r^2 = 0.73$ in group A and $r^2 = 0.67$ in Group B), combined with the low residual errors (RMSE = 0.055 g/cm^2 [6.3%] in group A and RMSE = 0.077 g/cm^2 [9.3%] in group B) further confirmed the actual strength of the relationship between O.S. values and DXA-measured BMD. These results are visually emphasized by the graph in Figure 7, containing the scatterplot of BMD values provided by the two techniques (DXA and US) for single patients in both age ranges. In Figure 8 the Bland–Altman plot is obtained, in which the overall average difference in BMD measurement (expressed as bias ± 2 SDs) was 0.003 ± 0.141 g/cm^2 and the corresponding values for single age ranges were 0.005 ± 0.127 g/cm^2 for group A and 0.002 ± 0.155 g/cm^2 for group B.

DISCUSSION

This study preliminarily illustrated the feasibility of a novel US technique for the diagnosis of osteoporosis that is applicable to the lumbar spine and provides a new numerical parameter, called the osteoporosis score, which directly correlates with DXA BMD measurements and consequently with DXA diagnostic evaluations.

As a first step, a reference database was built and a pair of model spectra (a "healthy" one and an "osteoporotic" one) was calculated for each age interval considered, optimizing the choice of the patients to be included in each model to maximize diagnostic accuracy.

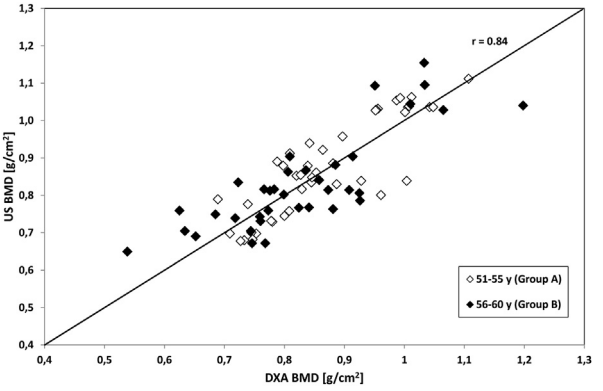

Fig. 7. Scatterplot of bone mineral density (BMD) values provided by dual X-ray absorptiometry (DXA) and ultrasound measurements for patients belonging to the study population. The line of equality and the global Pearson correlation coefficient are also illustrated.

If it is assumed that DXA is the gold standard reference, US-based identification of osteoporotic, osteopenic and healthy patients employing the selected model spectra, and the corresponding O.S. diagnostic thresholds, was correct for 95.7% of the patients included in the reference database and not used for model calculation (n = 188, acquisitions performed by an experienced operator). A similar level of accuracy (91.1%) (Table 5) was obtained on study population patients whose acquisitions were performed by inexperienced operators (n = 79). In both cases, agreement between the two diagnostic methods was substantially and significantly higher than what would be expected on the basis of pure chance, as documented by Cohen's κ (Tables 2 and 5). Another important aspect was the fact that there were no cases

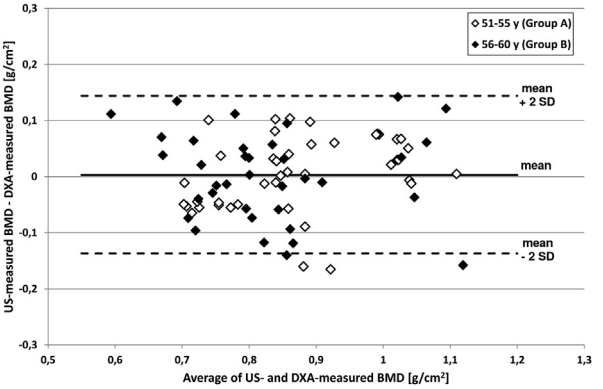

Fig. 8. Bland–Altman plot for comparison of ultrasound (US)- and dual X-ray absorptiometry (DXA)-measured bone mineral density (BMD) values for patients belonging to the study population.

in which "osteoporotic" patients were erroneously classified as "healthy" or vice versa.

Study population data sets were also used to derive US-based estimates of BMD values, which were found to be highly correlated with the corresponding DXA measurements: We observed coefficients of determination (r^2) in the range 0.67–0.73 with standard errors of estimate of 6.3% to 9.3%, indicating that the results obtained are robust and can be obtained using the same approach in independent samples. These results, obtained on a sample of 79 women, are comparable to those reported by Barkmann et al. (2010) employing a QUS approach to femoral BMD measurement; their sample comprised 62 women and they obtained an r^2 of 0.72 (RMSE ~10%) between DXA results and QUS-based estimates. On the other hand, the only previous study involving the clinical measurement of spinal BMD using an US method is that of Garra et al. (2009) in which the number of patients was small (n = 9) and the reported correlation with DXA-measured BMD was moderate (r^2 ~0.37).

The peculiar feature of our method is the exploitation of RF signals acquired during an echographic scan of the target bone structure to determine whether the internal bone architecture can be labeled as "osteoporotic" or "healthy" through detailed comparison with reference spectral models.

It is interesting to observe that, in our case, the discrepancies between US and DXA results were not necessarily a result of a lower accuracy of the adopted US methodology, but could be at least partially attributed to possible DXA inaccuracies. For instance, results of the adopted US method, unlike DXA, are not affected by aortic wall calcifications, which are automatically excluded by the algorithm during the vertebra identification step.

Another important feature of the proposed method is its full automation, which reduces dependence on operator experience to a minimum. In fact, as documented earlier, the implemented algorithm automatically identifies and discards "noisy" acquisitions, ensuring that diagnostic evaluations are performed only on US data sets reaching a specifically determined quality threshold. Our study also illustrated the extreme ease of use of the described system, given that inexperienced operators were able to perform US scans of suitable quality in 96.3% of cases, and the potential compliance of the adopted protocol with time constraints of clinical routine, as each patient could be examined and diagnosed in 3 min (1 min for US scan and 2 min for automatic data analysis).

Commercially available QUS devices for osteoporosis diagnosis are currently applicable only on peripheral bone regions, and their recognized value is actually limited to fragility fracture prediction in patients >65 y through calcaneal measurements, whose outcome has to be employed in conjunction with clinical risk factors (ISCD 2013). One of the main limitations of peripheral QUS systems is their poor or moderate correlation with spinal DXA outputs (Dane et al. 2008; El Maghraoui et al. 2009; Gemalmaz et al. 2007; Iida et al. 2010; Kwok et al. 2012). However, as expected, the adoption of an innovative US approach exploiting site-matched vertebral measurements resulted in markedly improved correlation. In fact, by reviewing literature quantifying the correlation of current peripheral QUS techniques with spinal DXA results through clinical studies in cohorts of women, we found that r^2 was always <0.38 (apart from the very recent article by Jiang et al. [2014], who adopted an experimental backscatter technique for calcaneal measurements and obtained r^2 = 0.56 with DXA-measured spinal BMD), emphasizing the value of our reported results. We also documented a measurement precision (RMS-CV = 2.95%) that is comparable to the typical values reported for clinically available peripheral QUS devices (Njeh et al. 2000), but is coupled with the aforementioned higher accuracies.

It is important to note that the US signal portions used in this study for spectral model constructions and O.S. value calculations are essentially related to the trabecular part of vertebrae: The development of an extended data analysis protocol, capable of taking into account cortical properties as well, could provide even better correlations with DXA-measured BMD, as preliminarily illustrated by a very recent pilot study focused on "*ex vivo*" QUS assessment of femoral strength (Grimal et al. 2013). Furthermore, one should consider that even if fracture discrimination was not explicitly involved in the present study, given the significant correlations obtained between O.S.-derived BMD values and the corresponding data provided by DXA, currently representing the gold standard technique for the estimation of bone fragility and fracture risk, our approach can reasonably be expected to perform similarly to DXA in fracture risk assessment as well.

Moreover, it is worth observing that the highly selective automatic identification of vertebrae and related ROIs, combined with the significant statistical basis of our proposed approach (requiring the described series of averaging and normalization operations on signals and spectra), has the potential to at least partially overcome random interference noise, one factor limiting the precision of US backscatter measurements. Image and signal selections and the sequences of averaging operations reduce the incidence of any kind of random effect and are also an indirect way to take into account, as a first approximation, that US velocity can vary between different vertebrae and different patients. Actually, our proposed approach, in its present implementation, differs

from previously reported approaches because it is based on overall correlations between different spectra, each considered as a whole without extracting any synthetic parameter and without associating a specific meaning to single spectrum peaks or valleys. All the local characteristics of the considered spectra are indirectly taken into account by the illustrated correlation process, intrinsically providing a more statistically significant basis of the reference data analysis. This, coupled with the statistical derivation of reference models starting from real human data, is probably the reason for the improvement in the correlation between DXA-measured BMD and O.S.-based estimates with respect to different US parameters reported in the literature (*e.g.*, spectral centroid shift), involving only specific spectral features whose values are typically compared with phantom measurements.

With respect to the theoretical interpretation of our reference models, we can adapt the approach introduced by Hoffmeister et al. (2012) for calculating the backscatter difference spectrum between two signal portions gated at different depths on the same signal backscattered from a cancellous bone sample. In our case, for each couple of spectral models, the osteoporotic model is considered as the spectrum of the first gated signal (inferior depth) and the healthy model as the spectrum of the second gated signal (superior depth).

The difference spectra obtained are illustrated in Figure 9, which also indicates the analysis bandwidth calculated using the procedure reported by Hoffmeister et al. (2012), corresponding to 1.5–2.8 MHz. With reasonable approximation, both difference spectra were monotonically increasing quasi-linear functions of frequency, similar to those reported in the article cited, providing a rough indication of the suitability of the assumptions made, especially in the case of group B models.

By considering the two curves in Figure 9 with respect to the whole frequency interval 1–5 MHz, we

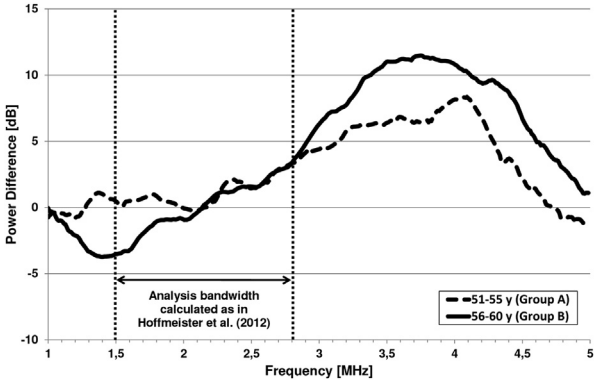

Fig. 9. Difference spectra derived for each pair of the described reference models.

can observe that in the case of group A models, the difference spectrum is about zero in the range 1–2 MHz (*i.e.*, the two models are almost coincident); then it linearly increases up to a peak at about 4 MHz and finally decreases, being again close to zero in the range 4.7–5 MHz; on the other hand, the difference spectrum for group B models had somewhat different behavior, being characterized by a kind of 'S' shape with a negative peak at about 1.4 MHz followed by a continuous increase up to a broad peak centered at about 3.8 MHz and a subsequent decrease until 5 MHz. Therefore, for patients in group A, the O.S. values are essentially due to spectral differences in the range 2–4.7 MHz, whereas for patients in group B, the two models show measurable differences that can contribute to the O.S. values in almost the whole range 1–5 MHz. This can be explained by hypothesizing that for the younger patients considered (group A) the spectral alterations caused by osteoporosis are evident only in the right part of the spectrum (*i.e.*, after the peak), whereas for the older patients (group B), such alterations become more evident and also start affecting the left side of the spectrum (*i.e.*, before the peak; see Fig. 6).

Finally, in analogy to the difference spectrum from two portions of the same backscattered signal, the fact that in a wide frequency range (approximately 2.8–5 MHz) the difference spectrum of the models in group B is shifted upward (to greater power differences) would correspond to an increase in temporal distance between the two gated signals (Hoffmeister et al. 2012), whereas in our case, it can be related to the age range of the patients considered. However, the fact that both difference spectra illustrated in Figure 9 feature a peak at a similar frequency (3.8 MHz vs. 4 MHz), although presenting different trends, is a useful indication for possible method improvements based on the introduction of "weighted correlations" aimed at better exploiting the frequency ranges in which the shape differences between the models are increased.

A more detailed analysis of the effects of single spectral features on the final O.S. value, with theoretical modeling of the related phenomena, will require further investigation based on the development of more complex experimental models. This is actually the subject of our ongoing studies, involving the *in vitro* application of the described technique on excised human bone samples, which are analyzed in a laboratory setup with all the boundary conditions known and the results of which are compared with quantitative micro-computed tomography measurements.

Study limitations

Limitations of this study are the mainly experimental and statistical nature of the proposed approach,

the adoption of DXA as gold standard reference, the number of clinical centers involved and the enrollment criteria.

The adopted method is essentially on statistical considerations and on the interpretation of experimental data, and further refinements would be required to remove possible artifacts that could arise from the analysis of different patient populations: (i) The fact that US velocity differs in different vertebrae and from one person to another was taken into account only indirectly. (ii) The observed spectral changes could be partially caused by US propagation through soft tissue. (iii) The key spectral features affecting O.S. values were not investigated in detail in the present study; we relied on previous scientific referenced articles to provide an initial theoretical interpretation of the results obtained. On one hand, the protocol employed is supported by numerous statistical considerations, and this should contribute to strengthening the significance of the results obtained and to overcoming the main current limitation of US backscatter measurements, which is random interference noise. On the other hand, although a detailed theoretical modeling of reported results was beyond the scope of the present work, we note that in several points, our discussed results are in agreement with previously published articles on US backscatter measurements for osteoporosis diagnosis (*e.g.*, healthier bones act as a kind of low-pass filter on the backscattered signals; US backscatter measurements can provide appreciable correlations with BMD and bone quantity parameters in general).

The use of DXA as a gold standard reference was not strictly correct because DXA represents a recognized gold standard for BMD measurements, but does not provide any information on bone quality and strength, which can affect US measurements. Therefore, the employment of quantitative computed tomography as an additional or alternative term of comparison to evaluate US measurement results would be more appropriate. However, this also implies that in some cases, the discrepancy between DXA and US measurements could be not attributed entirely to an US error, but (at least partially) to a DXA misevaluation.

Involvement of a single clinical center and recruitment based on referral for a spinal DXA investigation independently from motivation obviously influenced the study population characteristics, although representing the real field of employment of current systems for osteoporosis diagnosis.

Finally, the implemented approach can in principle be applied to all bone regions, without respect to patient age, sex and ethnic group, but the present study involved only spine investigations on Caucasian women in the age range 51–60 y with a BMI <25 kg/m^2. Therefore, further studies on different patient cohorts are necessary to determine the general validity of the method, although the spine is clearly the most difficult bone region to be reached with US, and moreover, there are no real reasons to expect significant variations in accuracy as a function of sex and/or ethnic group, provided that the appropriate reference databases for suitable model calculations are established. Increased BMI values could, in theory, be a challenge because of the augmented thickness of soft tissue between skin and vertebrae, but this issue should be successfully overcome through additional tailored optimization of the parameters involved in US data acquisition (*i.e.*, echograph power, focus position, scan depth, time gain compensation, amplification gain).

CONCLUSIONS

A novel US-based method for spinal densitometry was described; its clinical feasibility was illustrated and its diagnostic agreement with DXA evaluations documented. Because of its accuracy levels combined with the complete absence of ionizing radiation and its proven ease of use, this method has potential in future applications for diagnosis of osteoporotic disease at an earlier stage through population mass screenings. Moreover, the US assessment of internal bone structure described not only provides information related to BMD, but could also provide further information on the structural quality of bone and its real strength, offering exciting new perspectives regarding direct and accurate prediction of fracture risk. Future studies will include wider multicenter clinical validations involving additional bone regions (*e.g.*, proximal femur) and additional gold standard references (*e.g.*, quantitative computed tomography), with the specific aim of identifying single indicators of bone status (BMD, quality, strength) that actually contribute to O.S. value definition.

Acknowledgments—This work was partially funded by Programma Operativo Fondo Europeo di Sviluppo Regionale, Apulia Region 2007–2013, Action 1.2.4 (Grant N. 3Q5AX31: ECHOLIGHT Project).

REFERENCES

Altman DG, Bland JM. Measurements in medicine: The analysis of method comparison studies. Statistician 1983;32:307–317.

Baim S, Leslie WD. Assessment of fracture risk. Curr Osteoporos Rep 2012;10:28–41.

Barkmann R, Dencks S, Laugier P, Padilla F, Brixen K, Ryg J, Seekamp A, Mahlke L, Bremer A, Heller M, Gluer CC. Femur ultrasound (FemUS)—First clinical results on hip fracture discrimination and estimation of femoral BMD. Osteoporos Int 2010;21:969–976.

Barkmann R, Laugier P, Moser U, Dencks S, Klausner M, Padilla F, Haiat G, Gluer CC. A device for in vivo measurements of quantitative ultrasound variables at the human proximal femur. IEEE Trans Ultrason Ferroelectr Freq Control 2008a;55:1197–1204.

Barkmann R, Laugier P, Moser U, Dencks S, Klausner M, Padilla F, Haiat G, Heller M, Gluer CC. In vivo measurements of ultrasound transmission through the human proximal femur. Ultrasound Med Biol 2008b;34:1186–1190.

Barkmann R, Laugier P, Moser U, Dencks S, Padilla F, Haiat G, Heller M, Gluer CC. A method for the estimation of femoral bone mineral density from variables of ultrasound transmission through the human femur. Bone 2007;40:37–44.

Botella S, Restituto P, Monreal I, Colina I, Calleja A, Varo N. Traditional and novel bone remodeling markers in premenopausal and postmenopausal women. J Clin Endocrinol Metab 2013;98:E1740–E1748.

Brambilla M, De Mauri A, Leva L, Carriero A, Picano E. Cumulative radiation dose from medical imaging in chronic adult patients. Am J Med 2013;126:480–486.

Breban S, Padilla F, Fujisawa Y, Mano I, Matsukawa M, Benhamou CL, Otani T, Laugier P, Chappard C. Trabecular and cortical bone separately assessed at radius with a new ultrasound device, in a young adult population with various physical activities. Bone 2010;46:1620–1625.

Byberg L, Gedeborg R, Cars T, Sundstrom J, Berglund L, Kilander L, Melhus H, Michaelsson K. Prediction of fracture risk in men: A cohort study. J Bone Miner Res 2012;27:797–807.

Chanchairujira K, Chung CB, Kim JY, Papakonstantinou O, Lee MH, Clopton P, Resnick D. Intervertebral disk calcification of the spine in an elderly population: Radiographic prevalence, location, and distribution and correlation with spinal degeneration. Radiology 2004;230:499–503.

Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:37–46.

Cummings SR, Black DM, Nevitt MC, Browner W, Cauley J, Ensrud K, Genant HK, Palermo L, Scott J, Vogt TM. The Study of Osteoporotic Fractures Research Group. Bone density at various sites for prediction of hip fractures. Lancet 1993;341:72–75.

Cummings SR, Cawthon PM, Ensrud KE, Cauley JA, Fink HA, Orwoll ES. BMD and risk of hip and nonvertebral fractures in older men: A prospective study and comparison with older women. J Bone Miner Res 2006;21:1550–1556.

Curtis JR, Safford MM. Management of osteoporosis among the elderly with other chronic medical conditions. Drugs Aging 2012;29:549–564.

Dane C, Dane B, Cetin A, Erginbas M. The role of quantitative ultrasound in predicting osteoporosis defined by dual-energy X-ray absorptiometry in pre- and postmenopausal women. Climacteric 2008;11:296–303.

De Laet C, Kanis JA, Oden A, Johanson H, Johnell O, Delmas P, Eisman JA, Kroger H, Fujiwara S, Garnero P, McCloskey EV, Mellstrom D, Melton LJ III, Meunier PJ, Pols HA, Reeve J, Silman A, Tenenhouse A. Body mass index as a predictor of fracture risk: A meta-analysis. Osteoporos Int 2005;16:1330–1338.

Dencks S, Barkmann R, Padilla F, Haiat G, Laugier P, Gluer CC. Wavelet-based signal processing of in vitro ultrasonic measurements at the proximal femur. Ultrasound Med Biol 2007;33:970–980.

Dencks S, Barkmann R, Padilla F, Laugier P, Schmitz G, Gluer CC. Model-based estimation of quantitative ultrasound variables at the proximal femur. IEEE Trans Ultrason Ferroelectr Freq Control 2008;55:1304–1315.

Edwards MH, Jameson K, Denison H, Harvey NC, Sayer AA, Dennison EM, Cooper C. Clinical risk factors, bone density and fall history in the prediction of incident fracture among men and women. Bone 2013;52:541–547.

El Maghraoui A, Morjane F, Mounach A, Ghazi M, Nouijai A, Achemlal L, Bezza A, Ghozlani I. Performance of calcaneus quantitative ultrasound and dual-energy X-ray absorptiometry in the discrimination of prevalent asymptomatic osteoporotic fractures in postmenopausal women. Rheumatol Int 2009;29:551–556.

Engelke K, Gluer CC. Quality and performance measures in bone densitometry: Part 1. Errors and diagnosis. Osteoporos Int 2006;17:1283–1292.

Ferrari S, Bianchi ML, Eisman JA, Foldes AJ, Adami S, Wahl DA, Stepan JJ, de Vernejoul MC, Kaufman JM. IOF Committee of Scientific Advisors Working Group on Osteoporosis Pathophysiology. Osteoporosis in young adults: Pathophysiology, diagnosis, and management. Osteoporos Int 2012;23:2735–2748.

Fleiss L. Statistical methods for rates and proportions. Hoboken, NJ: Wiley; 1981.

Garra BS, Locher M, Felker S, Wear KA. Measurements of ultrasonic backscatter red spectral centroid shift from spine in vivo: Methodology and preliminary results. Ultrasound Med Biol 2009;35:165–168.

Gemalmaz A, Discigil G, Sensoy N, Basak O. Identifying osteoporosis in a primary care setting with quantitative ultrasound: Relationship to anthropometric and lifestyle factors. J Bone Miner Metab 2007;25:184–192.

Grimal Q, Grondin J, Guerard S, Barkmann R, Engelke K, Gluer CC, Laugier P. Quantitative ultrasound of cortical bone in the femoral neck predicts femur strength: Results of a pilot study. J Bone Miner Res 2013;28:302–312.

Grondin J, Grimal Q, Engelke K, Laugier P. Potential of first arriving signal to assess cortical bone geometry at the hip with QUS: model based study. Ultrasound Med Biol 2010;36:656–666.

Haiat G, Padilla F, Barkmann R, Dencks S, Moser U, Gluer CC, Laugier P. Optimal prediction of bone mineral density with ultrasonic measurements in excised human femur. Calcif Tissue Int 2005;77:186–192.

Hoffmeister BK. Frequency dependence of apparent ultrasonic backscatter from human cancellous bone. Phys Med Biol 2011;56:667–683.

Hoffmeister BK, Johnson DP, Janeski JA, Keedy DA, Steiner BW, Viano AM, Kaste SC. Ultrasonic characterization of human cancellous bone in vitro using three different apparent backscatter parameters in the frequency range 0.6–15 MHz. IEEE Trans Ultrason Ferroelectr Freq Control 2008;55:1442–1452.

Hoffmeister BK, Wilson AR, Gilbert MJ, Sellers ME. A backscatter difference technique for ultrasonic bone assessment. J Acoust Soc Am 2012;132:4069–4076.

Hou YL, Liao EY, Wu XP, Peng YQ, Zhang H, Dai RC, Luo XH, Cao XZ. Effects of the sample size of reference population on determining BMD reference curve and peak BMD and diagnosing osteoporosis. Osteoporos Int 2008;19:71–78.

Iida T, Chikamura C, Aoi S, Ikeda H, Matsuda Y, Oguri Y, Ono Y, Katada K, Ishizaki F. A study on the validity of quantitative ultrasonic measurement used the bone mineral density values on dual-energy X-ray absorptiometry in young and in middle-aged or older women. Radiol Phys Technol 2010;3:113–119.

International Society for Clinical Densitometry (ISCD). Official positions of the ISCD as updated in 2013. Available at: http://www.iscd.org/official-positions/2013-iscd-official-positions-adult/. Accessed: July 29, 2013.

Jiang YQ, Liu CC, Li RY, Wang WP, Ding H, Qi Q, Ta D, Dong J, Wang WQ. Analysis of apparent integrated backscatter coefficient and backscatter spectral centroid shift in calcaneus in vivo for the ultrasonic evaluation of osteoporosis. Ultrasound Med Biol 2014;40:1307–1317.

Johansson H, Kanis JA, Odén A, McCloskey E, Chapurlat RD, Christiansen C, Cummings SR, Diez-Perez A, Eisman JA, Fujiwara S, Glüer CC, Goltzman D, Hans D, Khaw KT, Krieg MA, Kröger H, Lacroix AZ, Lau E, Leslie WD, Mellström D, Melton LJ III, O'Neill TW, Pasco JA, Prior JC, Reid DM, Rivadeneira F, van Staa T, Yoshimura N, Zillikens MC. A meta-analysis of the association of fracture risk and body mass index in women. J Bone Miner Res 2014;29:223–233.

Kanis JA, WHO Study Group. Assessment of fracture risk and its application to screening for postmenopausal osteoporosis: Synopsis of a WHO report. Osteoporos Int 1994;4:368–381.

Karjalainen JP, Riekkinen O, Toyras J, Hakulinen M, Kroger H, Rikkonen T, Salovaara K, Jurvelin JS. Multi-site bone ultrasound measurements in elderly women with and without previous hip fractures. Osteoporos Int 2012;23:1287–1295.

Karjalainen JP, Toyras J, Riekkinen O, Hakulinen M, Jurvelin JS. Ultrasound backscatter imaging provides frequency-dependent information on structure, composition and mechanical properties of human trabecular bone. Ultrasound Med Biol 2009;35:1376–1384.

Krieg MA, Barkmann R, Gonnelli S, Stewart A, Bauer DC, Del Rio Barquero L, Kaufman JJ, Lorenc R, Miller PD, Olszynski WP, Poiana C, Schott AM, Lewiecki EM, Hans D. Quantitative ultrasound in the management of osteoporosis: The 2007 ISCD Official Positions. J Clin Densitom 2008;11:163–187.

Kwok T, Khoo CC, Leung J, Kwok A, Qin L, Woo J, Leung PC. Predictive values of calcaneal quantitative ultrasound and dual energy X-ray absorptiometry for non-vertebral fracture in older men: Results from the MrOS study (Hong Kong). Osteoporos Int 2012;23:1001–1006.

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

Lewiecki EM. Bone densitometry and vertebral fracture assessment. Curr Osteoporos Rep 2010;8:123–130.

Link TM. Osteoporosis imaging: State of the art and advanced imaging. Radiology 2012;263:3–17.

Liu JM, Ma LY, Bi YF, Xu Y, Huang Y, Xu M, Zhao HY, Sun LH, Tao B, Li XY, Wang WQ, Ning G. A population-based study examining calcaneus quantitative ultrasound and its optimal cut-points to discriminate osteoporotic fractures among 9352 Chinese women and men. J Clin Endorinol Metab 2012;97:800–809.

Liu XS, Shane E, McMahon DJ, Guo XE. Individual trabecula segmentation (ITS)-based morphologic analysis of microscale images of human tibial trabecular bone at limited spatial resolution. J Bone Miner Res 2011;26:2184–2193.

Looker AC, Wahner HW, Dunn WL, Calvo MS, Harris TB, Heyse SP, Johnston CC Jr, Lindsay R. Updated data on proximal femur bone mineral levels of US adults. Osteoporos Int 1998;8:468–489.

Moayyeri A, Adams JE, Adler RA, Krieg MA, Hans D, Compston J, Lewiecki EM. Quantitative ultrasound of the heel and fracture risk assessment: An updated meta-analysis. Osteoporos Int 2012;23:143–153.

Nayak S, Olkin I, Liu H, Grabe M, Gould MK, Allen IE, Owens DK, Bravata DM. Meta-analysis: Accuracy of quantitative ultrasound for identifying patients with osteoporosis. Ann Intern Med 2006;144:832–841.

Njeh CF, Hans D, Li J, Fan B, Fuerst T, He YQ, Tsuda-Futami E, Lu Y, Wu CY, Genant HK. Comparison of six calcaneal quantitative ultrasound devices: Precision and hip fracture discrimination. Osteoporos Int 2000;11:1051–1062.

O'Donnell M, Miller JG. Quantitative broadband ultrasonic backscatter: An approach to non-destructive evaluation in acoustically inhomogeneous materials. J Appl Phys 1981;52:1056–1065.

Padilla F, Jenson F, Bousson V, Peyrin F, Laugier P. Relationships of trabecular bone structure with quantitative ultrasound parameters: In vitro study on human proximal femur using transmission and backscatter measurements. Bone 2008;42:1193–1202.

Paggiosi MA, Barkmann R, Gluer CC, Roux C, Reid DM, Felsenberg D, Bradburn M, Eastell R. A European multicenter comparison of quantitative ultrasound measurement variables: The OPUS study. Osteoporos Int 2012;23:2815–2828.

Pais R, Campean R, Simon SP, Bolosiu CR, Muntean L, Bolosiu HD. Accuracy of quantitative ultrasound parameters in the diagnosis of osteoporosis. Centr Eur J Med 2010;5:478–485.

Pedrazzoni M, Girasole G, Bertoldo F, Bianchi G, Cepollaro C, Del Puente A, Giannini S, Gonnelli S, Maggio D, Marcocci C, Minisola S, Palummeri E, Rossini M, Sartori L, Sinigaglia L. Definition of a population-specific DXA reference standard in Italian women: The Densitometric Italian Normative Study (DINS). Osteoporos Int 2003;14:978–982.

Picano E, Matucci-Cerinic M. Unnecessary radiation exposure from medical imaging in the rheumatology patient. Rheumatology 2011;50:1537–1539.

Picano E, Vano E. Radiation exposure as an occupational hazard. Eurointervention 2012;8:649–653.

Pike C, Birnbaum HG, Schiller M, Sharma H, Burge R, Edgell ET. Direct and indirect costs of non-vertebral fracture patients with osteoporosis in the US. Pharmacoeconomics 2010;28:395–409.

Rand T, Seidl G, Kainberger F, Resch A, Hittmair K, Schneider B, Glüer CC, Imhof H. Impact of spinal degenerative changes on the evaluation of bone mineral density with dual energy X-ray absorptiometry (DXA). Calcif Tissue Int 1997;60:430–433.

Roux C, Roberjot V, Porcher R, Kolta S, Dougados M, Laugier P. Ultrasonic backscatter and transmission parameters at the os calcis in postmenopausal osteoporosis. J Bone Miner Res 2001;16:1353–1362.

Sambrook P, Cooper C. Osteoporosis. Lancet 2006;367:2010–2018.

Schnitzer TJ, Wysocki N, Barkema D, Griffith J, Lent V, Romba M, Welbel R, Bhuva S, Manyam B, Linn S. Calcaneal quantitative ultrasound compared with hip and femoral neck dual-energy X-ray absorptiometry in people with a spinal cord injury. PM R 2012;4:748–755.

Semelka RC, Armao DM, Elias J Jr, Picano E. The information imperative: Is it time for an informed consent process explaining the risks of medical radiation? Radiology 2012;262:15–18.

Sigelmann RA, Reid JM. Analysis and measurement of ultrasound backscattering from an ensemble of scatterers excited by sine-wave bursts. J Acoust Soc Am 1973;53:1351–1355.

Stewart A, Felsenberg D, Eastell R, Roux C, Glüer CC, Reid DM. Relationship between risk factors and QUS in a European population: The OPUS study. Bone 2006;39:609–615.

Szklarska A, Lipowicz A. BMI, hypertension and low bone mineral density in adult men and women. Homo 2012;63:282–291.

Trimpou P, Bosaeus I, Bengtsson BA, Landin-Wilhelmsen K. High correlation between quantitative ultrasound and DXA during 7 y of follow-up. Eur J Radiol 2010;73:360–364.

Van den Bergh JP, van Geel TA, Geusens PP. Osteoporosis, frailty and fracture: Implications for case finding and therapy. Nat Rev Rheumatol 2012;8:163–172.

Wear KA. The effect of phase cancellation on estimates of broadband ultrasound attenuation and backscatter coefficient in human calcaneus in vitro. IEEE Trans Ultrason Ferroelectr Freq Control 2008;55:384–390.

Wear KA, Nagaraja S, Dreher M, Gibson SL. Relationships of quantitative ultrasound parameters with cancellous bone microstructure in human calcaneus in vitro. J Acoust Soc Am 2012;131:1605–1612.

APPENDIX: DETAILED SEQUENCE OF THE PROCESSING STEPS PERFORMED ON EACH DATA FRAME ACQUIRED ON THE *K*TH PATIENT CONSIDERED

1. Rearrangement of image data in a rectangular matrix: To simplify the subsequent processing steps, image data points corresponding to the single echographic scan lines are organized in a rectangular matrix called I of size $N_{pix} \times N_{lines}$, where N_{pix} is the number of samples per echographic line, and N_{lines} is the number of echographic lines (Fig. 4a, b).
2. Brightness masking: Image brightness is modified through apposite masks, which increase the brightness of the central region and attenuate brightness toward image boundaries (Fig. 4c, d). In more detail, first a power-like compensation designed to attenuate the upper and lower image portions is applied; each element $I[i,j]$ of the matrix I is raised to the power represented by the element $\mathrm{EXP}[i,j]$ of the matrix EXP, having $N_{pix} \times N_{lines}$ elements and being defined as

$$\mathrm{EXP}[i,j] = \begin{cases} \dfrac{i}{0.3 \cdot N_{pix}}; & i \le 0.3 \cdot N_{pix} \\ \dfrac{i + 0.2 \cdot N_{pix}}{0.5 \cdot N_{pix}}; & 0.3 \cdot N_{pix} < i \le 0.55 \cdot N_{pix} \\ 1.5; & 0.55 \cdot N_{pix} < i \le 0.65 \cdot N_{pix} \\ \dfrac{1.4 \cdot N_{pix} - i}{0.5 \cdot N_{pix}}; & 0.65 \cdot N_{pix} < i \le 0.9 \cdot N_{pix} \\ \dfrac{N_{pix} - i}{0.1 \cdot N_{pix}}; & 0.9 \cdot N_{pix} < i \le N_{pix} \end{cases} \quad \text{(A1)}$$

In this way, the matrix I_2 is obtained:

$$I_2[i,j] = I[i,j]^{\mathrm{EXP}[i,j]} \quad \text{(A2)}$$

The matrix I_2 is then multiplied element by element by two brightness masks (Mask_1 and Mask_2) that emphasize the central image portion along the horizontal and vertical directions, respectively:

$$I_3[i,j] = I_2[i,j] \cdot \mathrm{Mask}_1[i,j] \tag{A3}$$

$$I_4[i,j] = I_3[i,j] \cdot \mathrm{Mask}_2[i,j] \tag{A4}$$

where Mask_1 and Mask_2 have $N_{\mathrm{pix}} \times N_{\mathrm{lines}}$ elements and are defined as

$$\mathrm{Mask}_1[i,j] = \begin{cases} \dfrac{2j+N_{\mathrm{lines}}-4}{2\cdot N_{\mathrm{lines}}-4}; & 1 \le j \le \dfrac{N_{\mathrm{lines}}}{2} \\ \dfrac{3\cdot N_{\mathrm{lines}}-2-2j}{2\cdot N_{\mathrm{lines}}-4}; & \dfrac{N_{\mathrm{lines}}}{2}+1 \le j \le N_{\mathrm{lines}} \end{cases} \tag{A5}$$

$$\mathrm{Mask}_2[i,j] = \begin{cases} 0.5; & i \le 0.18\cdot N_{\mathrm{pix}} \\ \dfrac{i+0.14\cdot N_{\mathrm{pix}}}{0.64\cdot N_{\mathrm{pix}}}; & 0.18\cdot N_{\mathrm{pix}} < i \le 0.5\cdot N_{\mathrm{pix}} \\ 1; & 0.5 < i \le 0.6\cdot N_{\mathrm{pix}} \\ \dfrac{1.2\cdot N_{\mathrm{pix}}-i}{0.6\cdot N_{\mathrm{pix}}}; & 0.6\cdot N_{\mathrm{pix}} < i \le 0.9\cdot N_{\mathrm{pix}} \\ 0.5; & 0.9\cdot N_{\mathrm{pix}} < i \le N_{\mathrm{pix}} \end{cases} \tag{A6}$$

3. Contrast enhancement and image smoothing: Pixel values (*i.e.*, values of I_4 elements) are normalized in the range between "0" and "1" and the following filtering sequence is applied (Fig. 4e–f):
   a. Contrast-limited adaptive histogram equalization. The image is divided into 64 identical rectangular regions, called *tiles*; each tile's histogram is equalized; and the neighboring tiles are then combined using bilinear interpolation to eliminate artificially induced boundaries.
   b. Two-dimensional low-pass Gaussian filter (size = 100 × 100, SD = 10).
   c. Additional "contrast-limited adaptive histogram equalization."
4. Histogram equalization on the whole image (Fig. 4g, h), which is necessary to make the results of the next step more robust.
5. Thresholding: Image is thresholded to become a binary map (threshold value = 0.985) (Fig. 4i, j).
6. Morphologic evaluations: All the clusters of white pixels identified in the thresholded image are automatically measured, and those with a length outside the range 20–45 mm are excluded from subsequent evaluations (see clusters marked with 'X' in Fig. 4k, l). If all the clusters are excluded, the frame is discarded and a new frame is considered (for instance, this is the case of the frame in Fig. 4l); otherwise, if there are retained clusters (*e.g.,* clusters labeled as 1, 2 and 3 in Fig. 4k), these clusters are ranked according to the following criteria:
   a. Length: Two points are assigned to the longest cluster and one point to the second one.
   b. Lateral position: Two points are assigned to the cluster that is the closest to the image center along the horizontal direction and one point to the second one.
   c. Vertical position: Two points are assigned to the cluster that is the closest to the image center along the vertical direction and one point to the second one.
   d. Average vertical thickness: Two points are assigned to the thinnest cluster and one point to the second one.

   The cluster with the highest score is labeled as a "possible vertebral interface" and passes to the subsequent spectral validation (in the case of Fig. 4k, cluster 1 received 2 points, cluster 2 received 6 points and cluster 3 received 4 points; therefore cluster "2" was labeled as a "possible vertebral interface").
7. Spectral validation: Once a "possible vertebral interface" has been found on the image, the algorithm also analyzes the corresponding RF data of a ROI selected as in the case of model construction (200-point Hamming-windowed signal portions, starting after the echo from the vertebral surface when the amplitude of RF signal envelope reached 15% of its peak value, and zero-padded to 4096 points). A comparative analysis is performed on the corresponding FFT power spectra, which are compensated and normalized as previously described to put them in the form $P_{\mathrm{Norm}_{ij}}(f)$ (where i identifies the considered vertebra and j indicates the specific echographic line). The "possible vertebral interface" is finally labeled as an "actual vertebral interface" if at least 70% of the spectra of the identified ROI have $r \geq 0.85$ with at least one of the appropriate reference model spectra (*i.e.*, the age-matched "healthy" and "osteoporotic" models, $\overline{\mathrm{MS}}_{\mathrm{Heal}_{\overline{N_y}}}(f)$ and $\overline{\mathrm{MS}}_{\mathrm{Ost}_{\overline{N_y}}}(f)$).
8. If an "actual vertebral interface" has been identified in the previous step, the subsequent three frames are skipped without analyzing them, to avoid reconsidering similar views of the same vertebra; otherwise the considered frame is discarded and no skip is performed.
9. A new analysis is started on the subsequent frame.