Evaluation of bone mineralization in former preterm born children: Phalangeal quantitative ultrasound cannot replace dual-energy X-ray absorptiometry

Carmen M.T. Lageweg, Mayke E. van der Putten, Johannes B. van Goudoever, Ton Feuth, Martin Gotthardt, Arno F.J. van Heijst, Viola Christmann

Keywords: Bone development, Children, Correlation, Dual-energy X-ray absorptiometry-scan, Speed of sound, Ultrasound

Abstract

Background: Preterm infants are at risk of impaired bone health in later life. Dual-energy X-ray absorptiometry-scan (DXA) is the gold standard to determine bone mineralization. Phalangeal quantitative ultrasound (pQUS) is an alternative technique that is inexpensive, easy to use and radiation-free. The aim of this study was to investigate whether both techniques reveal equivalent results.

Materials and methods: Sixty former preterm infants (31 boys; 29 girls) received a DXA and pQUS at age 9 to 10 years. DXA measured bone mineral content (BMC) and bone mineral density (BMD) for total body and lumbar spine (L1-4), while pQUS measured the amplitude dependent speed of sound (AD-SoS) and bone transit time (BTT) at metacarpals II-IV providing continuous values and Z-scores based on age and sex. Four statistical methods evaluated the association between both techniques: Pearson's correlation coefficients, partial correlation coefficients adjusted for gestational age, height and BMI, Bland-Altman analysis and cross tabulation.

Results: Both techniques showed a statistically significant weak correlation for continuous values as well as Z-scores (0.291–0.462, p < 0.05). Boys had significant and relatively high correlations (0.468–0.585, p < 0.05). In comparison, the correlations for girls were not significant. Correlation coefficients further decreased while calculating the partial correlations. The Bland-Altman plots showed poor agreement. Sensitivity ranged from 33% to 92% and specificity from 16% to 68%. Positive and negative predictive values ranged from 4% to 38% and 82% to 97%, respectively.

Conclusions: We found statistically significant weak correlations and poor agreement between DXA and pQUS measurements. DXA is not equivalent to pQUS and therefore not replaceable by this technique in former preterm born children at the age of 9 to 10 years.

1. Introduction

Bone development is one of the key processes during fetal, neonatal and infant development (Schoenau et al., 2004). Mineralization of bone mainly starts during the third trimester of pregnancy based on active placental transfer of calcium and phosphorus to the fetus. Up to 80% of the body calcium of a term infant is accrued during the last trimester (Kovacs, 2014). Preterm infants miss out the active fetal bone development and therefore are at risk of reduced bone mineralization and development of osteopenia (Harrison et al., 2008). Inadequate bone mineralization is seen as a risk factor for the development of osteoporosis in later life, which is an important cause of morbidity and mortality in elderly people and a considerable factor of healthcare expenditure (Kannus et al., 1999; Javaid and Cooper, 2002; Leppälä et al., 2002).
The peak bone mass is attained before skeletal maturity (Bonjour et al., 1991). Any factor that influences the acquisition of peak bone mass may represent a mechanism to affect later osteoporosis risk. The evaluation of bone development in preterm children is relevant for the determination of the individual health risk as well as the evaluation of medical treatment that aimed at improvement in bone development.

Currently, there are two techniques available to determine bone mineralization, either dual-energy X-ray absorptiometry-scan (DXA) or quantitative ultrasound (QUS). DXA is the most commonly used technique for assessing bone mineralization in children and adolescents (Wren and Gilsanz, 2006). Although DXA is a non-invasive and standardized method, it is not available for all medical centers and it uses a low amount of radiation. In recent years, QUS has been proposed as an alternative method to replace DXA for the evaluation of bone status, especially since it is relatively inexpensive, fast, easy to use, portable and radiation-free (Baroncelli, 2008; Gianni et al., 2008; Tansug et al., 2011; Tuna et al., 2008).

Studies investigating the association between the measurements of DXA and QUS revealed inconsistent results. While a number of studies showed a significant positive correlation between DXA and QUS (Van Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2006; Gonçalves et al., 2014; Hartman et al., 2004a; Pluskiewicz et al., 2002; Sani et al., 2011; Sundberg et al., 1998; Xu et al., 2014; Bjk-Drabik et al., 2016; Catalano et al., 2017; Oslynskja et al., 2016; Zuckerman-Levin et al., 2007; Mora et al., 2009; Weeks et al., 2016; Halaba et al., 2005; Catalano et al., 2013), others found a discrepancy between the measurements of the two methods (Gianni et al., 2008; Halaba et al., 2005; Chong et al., 2015; Christoforidis et al., 2010; Christoforidis et al., 2011; Hartman et al., 2004b; Williams et al., 2012; Alwis et al., 2010). This could be a result of the different QUS measurement sites or different patient categories investigated. Only a limited number of studies used the phalangeal QUS (Di Mase et al., 2012; Gonçalves et al., 2014; Pluskiewicz et al., 2002; Bjk-Drabik et al., 2016; Catalano et al., 2017; Oslynski et al., 2016; Halaba et al., 2005; Catalano et al., 2013) and only one study looked at the specific group of former preterm born children (Gianni et al., 2008).

The aim of this study was to investigate whether the measurements of dual-energy X-ray absorptiometry scan (DXA) and phalangeal quantitative ultrasound (pQUS) performed in preterm born children aging from 9 to 10 years reveal comparable results. We hypothesized that both techniques were equivalent in diagnosing the state of bone mineralization. Equivalent results would mean that the pQUS could replace the DXA for evaluation bone mineralization as a diagnostic tool.

2. Materials and methods

2.1. Study design

This study was a cross-sectional study using the data collection of the study “Long-term follow up of growth and bone mineralization of former preterm infants” (FoBoMin). This study was approved by the Ethics committee (CMO nr 2013/594) of the Radboud University Medical Center. Informed consent was obtained from all parents after approval by the local ethics committee.

2.2. Study population and procedure

The study included 60 former preterm infants at the age of 9 to 10 years. All subjects participated in the FoBoMin-study. This long-term follow-up study evaluated two cohorts of very preterm infants with a birth weight below 1500 g and gestational age < 34 weeks. The cohorts differed by nutritional intake during the first two weeks of life. The second cohort received higher intake of protein, energy as well as calcium and phosphate. This was associated with improved weight gain during the early postnatal period (Christmann et al., 2013). The aim of the FoBoMin-study was to compare long-term growth and bone mineralization in relation to early nutritional intake in preterm born children at age 9 to 10 years. All participants of the studies were evaluated by DXA and pQUS. The measurements were performed on the same day for the individual participant. Four statistical methods were used to compare both methods.

2.3. Measurement instruments and variables

Bone mineralization of the total body and lumbar spine (L1-L4) was determined using the QDR Discovery A S/N 85606 (Hologic, Inc., USA). According to the International Society for Clinical Densitometry (ISCD), the lumbar spine (L1-L4) and whole body scan are the preferred skeletal sites for measurement in children (Lewiecki et al., 2008). The measurements of the DXA were analyzed using the APEX-system software version 13.3. The DXA uses a low dose of radiation depending on measurement site. The effective dose, reflecting the real radiation risk for children of 10 years old, for the whole body is 4.8 μSv and for the lumbar spine 7.1 μSv (Blake et al., 2006). According to the ‘Rijksinstituut voor Volksgezondheid en Milieu’ (RIVM) the yearly averaged ambient dose equivalent rate for the NMR station in the area of Nijmegen is 74 mSv/h (Knetsch, 2013), resulting in a daily exposure in Nijmegen of 1.78 μSv. Therefore, the radiation dose of DXA can be regarded as very low and is negligible. Results of the DXA were expressed as Bone Mineral Content (BMC; g), Bone Mineral Density (BMD; g/cm²), representing the ratio between BMC and bone area (cm²), and Z-scores, representing the number of standard deviations above or below the mean for the patients’ sex and age. The Z-scores were calculated by the DXA software on the basis of reference values for sex and age obtained from a large U.S. population provided by the manufacturer. The Z-scores of the whole body were calculated using the reference data of the National Health and Nutrition Examination Survey (NHANES, 2008) (Kelly et al., 2009), while lumbar spine Z-scores were based on the reference data of the Bone Mineral Density in Childhood Study (BMDCS) (Zemel et al., 2011). A Z-score less than or equal to –2.0 SD is considered to indicate ‘low bone mineral status’ (Lewiecki et al., 2008).

The quantitative ultrasound (pQUS) was performed on the second to the fifth metacarpals of the phalangeal bones using a DBM Sonic Bone Profiler (IGEA, Carpi, Italy). The mean value of the measurements per person was calculated. The transmitter of the pQUS generated a sound frequency of 1.25 MHz. This technique measured the amplitude dependent speed of sound (AD-SoS) and bone transit time (BTT), which were both expressed in continuous values and in Z-scores. The AD-SoS (m/s) was the ultrasound velocity inside the finger and was derived from the measurement of the time interval between emission and reception of the ultrasound signal, considering the first signal with a minimum amplitude of 2 mV at the receiver probe. The BTT (μsec) reflected the bone characteristics without the interference of the soft tissue by calculating the difference between transmission time in soft tissue and bone and transmission time in soft tissue (Di Mase et al., 2012). The Z-scores were determined on the basis of the reference values related to sex and age (AD-SoS Z score (age); BTT- Z-score (age)) or sex and height (AD-SoS Z score (height); BTT Z-score (height)). The Z-scores were obtained from a large Italian population provided by the manufacturer (Baroncelli et al., 2006).

Additionally, age, sex, gestational age at birth, weight, height, BMI and pubertal development were recorded. Weight (kg) was measured using an electronic digital scale (SECA MOD701) to the nearest 0.1 kg. Height (cm) was measured using a vertical stadiometer (SECA MOD240) to the nearest 0.1 cm. Body mass index (BMI; kg/m²) was calculated by dividing weight (kg) by the square of height (m²). Pubertal development was self-assessed from pictures showing the different Tanner stages (Tye, 2016). The children were asked to indicate which picture most resembled their current appearance.
2.4. Statistical analysis

The statistical analysis was performed using the Statistical Package for the Social Sciences (IBM SPSS Inc., Chicago, IL, USA, version 22.0). All results were expressed as mean ± SD. Four statistical methods were used for the analysis of the association between pQUS and DXA. First, the Pearson’s correlation coefficients (r) were calculated for evaluation of the correlation between continuous values as well as the Z-scores of DXA and pQUS. The correlation coefficients were determined for every outcome for the total group as well as for boys and girls separately. Secondly, the partial correlation coefficients were determined to correct for possible confounders on the original correlation between DXA and pQUS. Possible confounders of bone development, such as age, sex, gestational age, weight, height, BMI and Tanner stages at follow-up were included in the analysis. Only three of these confounders, namely gestational age, height and BMI were used to calculate partial correlation coefficients, because of the limited number of participants in this study. The three confounders were chosen based on calculating whether they correlated significantly with DXA and pQUS measurements. Thirdly, a Bland-Altman analysis was performed to evaluate the agreement between both techniques using the Z-scores of either DXA and pQUS. Plots were created with the mean of two Z-scores within the same subject resulting from the two techniques on the horizontal axis and the difference of the Z-scores on the vertical axis. Finally, a cross tabulation was performed and the sensitivity, specificity as well as positive and negative predictive values were calculated, where DXA was considered as the gold standard. In agreement with the ISCD, a DXA Z-score less than or equal to −2.0 SDS should be considered as low bone mineralization (Lewiecki et al., 2008). A Z-score between −1.0 and −2.0 was considered as reduced (Acceto et al., 2014) and a Z-score above −1.0 SDS is normal. For the current study a cut-off value of −1.0 SDS was used for the assessment of low or normal bone mineralization, in the absence of participants with a Z-score less than −2.0.

A two tailed p-value of < 0.05 was considered statistically significant.

3. Results

3.1. Patient characteristics

The baseline characteristics of the participants (total group as well as boys and girls separately) are presented in Table 1. Anthropometric characteristics, gestational age at birth and pQUS measurements at follow-up were comparable between boys and girls. No statistically significant differences were found.

3.2. Correlation

Table 2 presents the correlation coefficients of DXA and pQUS measurements for the continuous values. The correlation coefficients between the DXA and both pQUS measurements (BTT; AD-SoS) showed statistical significance, though the r value was low. The correlation coefficients between DXA and BTT were higher, ranging from 0.341 to 0.462 (p < 0.05), compared to correlation coefficients between DXA and AD-SoS, ranging from 0.291 to 0.345 (p < 0.05). In comparison, boys showed a statistically significant and slightly higher correlation, which was not found for girls (boys: 0.468–0.585, p < 0.05 versus girls: 0.008–0.335, p > 0.05). Nevertheless, the differences found between boys and girls, calculated with the Fisher’s r-to-Z transformation, were not statistically significant, except for lumbar spine BMD and AD-SoS (p = 0.039).

Table 3 presents the correlation coefficients for the Z-scores. AD-SoS Z-score (age) and BTT Z-score (age) showed a statistically significant but weak correlation with DXA Z-scores (0.327–0.401, p ≤ 0.05). The correlation coefficients for BTT Z-scores (age) were higher than those for AD-SoS Z-scores (age). Since the Z-scores (height) showed no statistically significant correlation coefficients, they were not further evaluated. In comparison, the Z-scores (age) of boys showed a statistically significant correlation coefficient in contrast to girls (boys: 0.436–0.520, p < 0.05 versus girls: −0.026–0.274, p > 0.05). In general, the difference found between boys and girls, calculated with the Fisher’s r-to-Z transformation, was not statistically significant, except for lumbar spine Z-score and AD-SoS Z-score (age) (p = 0.027).

As an example, Fig. 1 illustrates an overlay scatterplot of the correlation coefficients between the AD-SoS Z-score (age) and the whole body Z-score for boys and girls. Other pQUS and DXA measurements revealed comparable scatterplots.

3.3. Partial correlation

Table 4 presents the original correlation coefficients and the partial correlation coefficients adjusted for gestational age, height and BMI for the continuous values and the Z-scores (age). The adjustment for the three confounders induced a further decrease of the correlation coefficients. On average, the remaining coefficients, although significant, were very weak.

3.4. Agreement

The agreement was considered for all parameters. Only the Bland-Altman plot of the whole body Z-score and the AD-SoS Z-score (age) will be presented, because the other plots showed comparable results (Fig. 2). The agreement between the two techniques was low, based on the following results. First, the mean difference between the Z-scores as presented in Fig. 2 was 2.73 and thereby significantly different from zero determined using a paired t-test (p-value: 0.011). Thereby, the 95%-limits of agreement in this figure had a large interval between −0.54 and 6.00. Lastly, the plot showed that the differences of the two Z-scores (whole body Z-score minus AD-SoS Z-score (age)) were negatively dependent on the mean. This means that the difference between DXA and pQUS Z-scores increased with a lower mean Z-score, leading to an increasing disagreement between the two techniques while assessing bone mineralization for lower Z-scores.

3.5. Cross tabulation

Cross tabulation was performed for all combinations of pQUS and DXA measurements. The number of pQUS measurements with a Z-score below −1.0 SDS was higher compared to DXA. The sensitivity for all measurements ranged from 33% to 92%. The specificity ranged from 16% to 68%. The positive and negative predictive values ranged from 4% to 38% and 82% to 97%, respectively.

As an example, Table 5 shows a cross table for the BTT Z-score (age) and the lumbar spine Z-score. This table revealed the best agreement of all measurements, but the specificity, sensitivity, positive predictive value and negative predictive value in general, were low. The sensitivity and specificity for BTT Z-score (age) in comparison with lumbar spine Z-score were 69% and 68%, respectively, and for the positive and negative predictive value this was 38% and 89%, respectively. Overall, we found a large discrepancy between the two methods for discriminating a patient with a normal or reduced bone mineralization.

4. Discussion

This study evaluated two different diagnostic techniques for bone development in former preterm born children, who are at risk for impaired bone mineralization. Four statistical tests showed that the results of dual-energy X-ray absorptiometry scan (DXA) and phalangeal quantitative ultrasound (pQUS) had a significant weak correlation that further decreased after adjustment for confounders. In addition, there was a low agreement between the two techniques and a discrepancy in differentiating the same children with normal or reduced bone
The association between QUS and DXA has previously been evaluated in our study of Halaba et al. (2005), who found a significant correlation between QUS and DXA. This is in agreement with the reference data provided by Di Mase et al. (2012; Falcini et al., 2000; Gonçalves Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2000; Gonçalves Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2000; Gonçalves Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2000; Gonçalves Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2000; Gonçalves Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2000; Gonçalves Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2000; Gonçalves Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2000; Gonçalves Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2000; Gonçalves Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2000; Gonçalves Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2000; Gonçalves Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2000; Gonçalves Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2000; Gonçalves Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2000; Gonçalves Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2000; Gonçalves Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2000; Gonçalves Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2000; Gonçalves Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2000; Gonçalves Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2000; Gonçalves Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2000; Gonçalves Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2000; Gonçalves Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2000; Gonçalves Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2000; Gonçalv

defined as a result of earlier skeletal maturation in girls compared to boys. Our children had a lower age range and were mainly prepubertal. We do not have an explanation for this phenomenon and therefore suggest these gender differences should be further investigated.

The association between QUS and DXA has previously been evaluated in a number of studies (Gianni et al., 2008; Tuna et al., 2008; Van Rijn et al., 2000; Di Mase et al., 2012; Falcini et al., 2000; Gonçalves et al., 2014; Hartman et al., 2004a; Pluskiewicz et al., 2002; Sani et al., 2011; Sundberg et al., 1998; Xu et al., 2014; Bajk-Drabik et al., 2016;...
SoS, amplitude-dependent speed of sound; BTT, bone transit time; pQUS, phalangeal quantitative ultrasound; DXA, dual-energy x-ray absorptiometry; AD-SoS, amplitude-dependent speed of sound; BTT, bone transit time; Z-score for sex and age; Z-score adjusted for sex and height.

All data are presented as correlation coefficients between pQUS and DXA measurements. Halaba et al. found no correlation for both sexes using the AD-SoS Z-score. Nevertheless, in agreement with our analyses these studies found comparable poor correlation coefficients and thereby questioning the equivalence of pQUS and DXA. In contrast, two other studies found relatively stronger correlations for the continuous values (0.59–0.74, p < 0.05) but no significant correlations while comparing Z-scores, except for the AD-SoS Z-score and whole body Z-score which had a poor significant correlation coefficient (0.31, p < 0.02) (Gonçalves et al., 2014; Baś-Drabik et al., 2016). These five studies evaluated children and adults at an age ranging from 4 to 27 years, while Catalano et al. (2013, 2017) evaluated postmenopausal women. The younger and smaller age range in our group could be an explanation for the different results we found.

Furthermore, Table 6 gives an overview of the authors’ conclusion per study, showing different interpretations for in general a relatively low correlation between QUS and DXA. An advantage of our study was that, besides the small range of age of the participants, more statistical tests were used to compare both techniques, leading to a more reliable overall conclusion.

The partial correlation coefficient was used to evaluate the effect of possible confounders on the original correlation. To our knowledge only a few studies that evaluated the association between pQUS and DXA looked at the influence of possible confounding factors with regard to correlation coefficients. Halaba et al. (2005) and Di Mase et al. (2012) performed a multiple linear regression analysis to evaluate the effect of anthropometric characteristics on DXA or QUS measurements using confounders such as age, sex, weight, height and BMI.

Only four studies used other statistical tests such as the Bland-Altman analysis (Chong et al., 2015; Williams et al., 2012) and cross tabulation (Tuna et al., 2008; Gonçalves et al., 2014; Chong et al., 2015; Williams et al., 2012). Although these studies looked at tibial and radial QUS, their conclusions are consistent with our findings of poor agreement between the two techniques and a large discrepancy in differentiating the same children as having normal or reduced bone mineralization.

The overall absence of an association between DXA and pQUS for continuous values and Z-scores could be explained by the fact that measurements were influenced by different bone composition and bone mineralization at different sites. In addition, children may be in different growth phases and some bones may grow faster than others, which potentially could have an effect on the bone development and thereby may have affected the results. Quantitative ultrasound can be applied on various parts of the extremities, such as the phalanx, radius, tibia and calcaneus. Recent studies suggest that the phalanges may be the most appropriate measurement site, because this site is sensitive to changes in bone status (Takada et al., 1997a; Ventura et al., 1996; Takada et al., 1997b).

For the assessment of pQUS we chose the DXA scan as the golden standard for comparison. Quantitative Computed Tomography (QCT), or peripheral QCT may have been good alternatives and may have provided more accurate results and additional information on bone strength (Pezzuti et al., 2017). However, QCT uses a relatively high radiation dose, especially for young children in research settings, and for both methods normative data for a pediatric population are lacking (Pezzuti et al., 2017; Fonseca et al., 2013). DXA has been recommended as an appropriate method for clinical densitometry of infants and young children by the International Society for Clinical Densitometry (ISCD) in 2013 (Gordon et al., 2013). The advantage of QUS is the ability to easily repeat measurements, giving the opportunity to follow the development of bone over time, especially under circumstances were diagnostic tools as DXA are not possible or available. Our results were limited to a single measurement per child. Therefore intra-individual repeated measurements with QUS should be further evaluated with regard to the reliability to predict the intra-individual development of bone over a longer time period.
Our study had several limitations. The number of patients was limited. Firstly, this limited the number of confounders we were able to use for calculation of the partial correlation coefficient. Secondly, low bone mineralization with a significant Z-score below 2.0 SD was only found in 2 out of 60 children for lumbar spine measurements, and none for the whole body measurements. To increase the number of children with deviant bone mineralization we chose a cut-off value of $-1.0$ SDS for determination of normal or low bone mineralization, in contrast to general practice. This is in accordance with the study of Gianni et al., who evaluated the prevalence of $Z$-scores $< -1.0$ SDS for tibial or radial QUS (Gianni et al., 2008), because the QUS values were higher in comparison with DXA. According to the literature a similar cut-off value can be used for AD-SoS and BTT (Di Mase et al., 2012; Baroncelli et al., 2006).

In general, the $Z$-scores were based on reference data derived from different populations. The reference data of the DXA $Z$-scores were based on an U.S. population, while the reference data for the pQUS $Z$-scores were recorded from an Italian population. It is not known whether the populations are comparable to our population and theoretically this might explain some of our differences.

5. Conclusion

This study demonstrated a weak association between DXA and pQUS measurements, established from a statistically significant weak correlation, a poor agreement and a discrepancy in differentiating the same children with normal or reduced bone mineralization. Therefore, pQUS measurements are not equivalent to DXA for the evaluation of bone health and cannot replace the DXA in former preterm born children at the age of 9 to 10 years.

Acknowledgements

The authors declare no conflict of interest. VC and AvH received a grant by Hero Kindervoeding, Breda, Netherlands and Fresenius Kabi Nederland B.V. Zeist. JbvG reports grants outside the submitted work from Mead Johnson Nutrition, during the conduct of the study, grants from EU, FP-7 programme. JbvG serves as director of the National Human Milk bank and as board member of the National health Council and the National Breastfeeding Committee.

CL and MvdP were responsible for data acquisition, analyzing and interpreting the data, and writing of the manuscript; TF supervised the statistical analyses and contributed to the interpretation of the data and review of the manuscript; MG supervised the design, interpretation of the data and reviewing the manuscript, JbvG, MG, and AvH contributed in the interpretation of the data and review of the manuscript; VC was responsible for the design, and supervised the analyses and interpretation of the data, and writing of the manuscript. All authors listed on the manuscript have seen and approved the manuscript and take full responsibility for the manuscript.
<table>
<thead>
<tr>
<th>Author (yr)</th>
<th>QUS</th>
<th>Age (yr)</th>
<th>Sample size</th>
<th>Morbidity</th>
<th>Statistical method</th>
<th>Outcome</th>
<th>Correlation coefficients (r)</th>
<th>Authors' conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pliskiewicz et al. (2002)</td>
<td>Phalanges</td>
<td>9-23</td>
<td>30</td>
<td>ESRF</td>
<td>LCC, SCC</td>
<td>AD-SoS, Z-score</td>
<td>0.45 - 0.56</td>
<td>+</td>
</tr>
<tr>
<td>Halaba et al. (2005)</td>
<td>Phalanges</td>
<td>14-19</td>
<td>150</td>
<td>Healthy</td>
<td>LRA, MRA</td>
<td>AD-SoS, TB, BMC</td>
<td>0.40-0.47, F/Z-scores: NS</td>
<td>−</td>
</tr>
<tr>
<td>Di Mase et al. (2012)</td>
<td>Phalanges</td>
<td>4-18</td>
<td>25</td>
<td>Idiopathic SH</td>
<td>PCC, MRA</td>
<td>AD-SoS, BTT, Z-score</td>
<td>0.42-0.52</td>
<td>+</td>
</tr>
<tr>
<td>Catalano et al. (2013)</td>
<td>Phalanges</td>
<td>43-78</td>
<td>80</td>
<td>Postmenopause</td>
<td>SCC</td>
<td>AD-SoS, BTT, UBIPI</td>
<td>FN BM/T, LS/BM</td>
<td>+</td>
</tr>
<tr>
<td>Gonçalves et al. (2014)</td>
<td>Phalanges</td>
<td>6-27</td>
<td>70</td>
<td>21-OHD</td>
<td>PCC, SCC, CT, ROC</td>
<td>AD-SoS, BTI, UBIPI, Z-score</td>
<td>0.59-0.72, Z-scores: no correlation</td>
<td>+/-</td>
</tr>
<tr>
<td>Baş-Drabik et al. (2016)</td>
<td>Phalanges</td>
<td>10-15</td>
<td>51</td>
<td>IBD</td>
<td>SCC</td>
<td>AD-SoS, TB, Z-score</td>
<td>Continuous: 0.69-0.74, TB Z-score: 0.31,</td>
<td>−</td>
</tr>
<tr>
<td>Catalano et al. (2017)</td>
<td>Phalanges</td>
<td>60-70</td>
<td>60</td>
<td>Postmenopause, BC</td>
<td>SCC, MRA</td>
<td>AD-SoS, BTI, UPI</td>
<td>0.21-0.29</td>
<td>+</td>
</tr>
<tr>
<td>Objaynki et al. (2016)</td>
<td>Radius/tibia/</td>
<td>30-97</td>
<td>4123</td>
<td>Healthy</td>
<td>PCC</td>
<td>SOS</td>
<td>0.58-0.66</td>
<td>+</td>
</tr>
<tr>
<td>Hartman et al. (2004a)</td>
<td>Radius/tibia/</td>
<td>4-18</td>
<td>40</td>
<td>CRD</td>
<td>PCC, MRA</td>
<td>SOS, Z-score</td>
<td>Radius: 0.54, Tibia: NS</td>
<td>+</td>
</tr>
<tr>
<td>Hartman et al. (2004b)</td>
<td>Radius/tibia/</td>
<td>5-18</td>
<td>41</td>
<td>Celiac disease</td>
<td>LRA, MRA</td>
<td>SOS, Z-score</td>
<td>0.3</td>
<td>−</td>
</tr>
<tr>
<td>Zuckerman-Levin et al. (2007)</td>
<td>Radius/tibia/</td>
<td>11-39</td>
<td>27</td>
<td>Turner's syndrome</td>
<td>SCC</td>
<td>SOS</td>
<td>0.46-0.63</td>
<td>−</td>
</tr>
<tr>
<td>Gianni et al. (2008)</td>
<td>Radius/tibia/</td>
<td>4-6</td>
<td>45</td>
<td>Ex-preterm</td>
<td>PCC</td>
<td>SOS, Z-score</td>
<td>NS</td>
<td>−</td>
</tr>
<tr>
<td>Mora et al. (2009)</td>
<td>Radius/tibia/</td>
<td>4-22</td>
<td>88</td>
<td>HIV-infection</td>
<td>MRA, x2 test</td>
<td>SOS, Z-score</td>
<td>0.58-0.66</td>
<td>+</td>
</tr>
<tr>
<td>Christoforidis et al. (2010)</td>
<td>Radius/tibia/</td>
<td>4-18</td>
<td>27</td>
<td>Hemophilia A</td>
<td>PCC, Cx</td>
<td>SOS, Z-score</td>
<td>Tibia: NS,_radius: -0.019</td>
<td>−</td>
</tr>
<tr>
<td>Christoforidis et al. (2011)</td>
<td>Radius/tibia/</td>
<td>2-16</td>
<td>20</td>
<td>CKD</td>
<td>PCC, SCC, MRA</td>
<td>SOS, Z-score</td>
<td>Tibia: 0.129, radius: 0.022</td>
<td>−</td>
</tr>
<tr>
<td>Williams et al. (2012)</td>
<td>Radius/tibia/</td>
<td>5-20</td>
<td>621</td>
<td>Healthy, CF, obese</td>
<td>CT, BAA</td>
<td>SOS, Z-score</td>
<td>0.39-0.44, Z-scores: NS</td>
<td>−</td>
</tr>
<tr>
<td>De Schepper et al. (2012)</td>
<td>Radius/</td>
<td>12-38</td>
<td>64</td>
<td>CF</td>
<td>LRA, MRA</td>
<td>SOS, Z-score</td>
<td>TB BMC/BM/Z-scores</td>
<td>−</td>
</tr>
<tr>
<td>Chong et al. (2015)</td>
<td>Radius</td>
<td>7-11</td>
<td>134</td>
<td>Healthy</td>
<td>CT, BAA, ROC</td>
<td>SOS, Z-score</td>
<td>0.63-0.81</td>
<td>+</td>
</tr>
<tr>
<td>Van Rijn et al. (2000)</td>
<td>Tibia</td>
<td>7-23</td>
<td>146</td>
<td>Healthy</td>
<td>MRA</td>
<td>SOS</td>
<td>0.29-0.36</td>
<td>−</td>
</tr>
<tr>
<td>Tuna et al. (2008)</td>
<td>Tibia</td>
<td>40-72</td>
<td>200</td>
<td>Postmenopause</td>
<td>LRA, SCC, CT, ROC</td>
<td>SOS</td>
<td>0.44-0.70</td>
<td>−</td>
</tr>
<tr>
<td>Sundberg et al. (1998)</td>
<td>Calcaneus</td>
<td>11-16</td>
<td>280</td>
<td>Healthy</td>
<td>PCC</td>
<td>SOS, BUA, SI</td>
<td>0.83</td>
<td>+</td>
</tr>
<tr>
<td>Falci et al. (2000)</td>
<td>Calcaneus</td>
<td>6-18</td>
<td>53</td>
<td>Juvenile arthritis</td>
<td>PCC, LRA</td>
<td>SOS, Z-score</td>
<td>BUA: 0.492-0.507, SOS: NS</td>
<td>−/+</td>
</tr>
<tr>
<td>Sani et al. (2011)</td>
<td>Calcaneus</td>
<td>6-17</td>
<td>8</td>
<td>Thalasemia</td>
<td>PCC</td>
<td>SOS, BUA, SI</td>
<td>0.690-0.693</td>
<td>+</td>
</tr>
<tr>
<td>Xu et al. (2014)</td>
<td>Calcaneus</td>
<td>5-19</td>
<td>392</td>
<td>Healthy</td>
<td>PCC</td>
<td>SOS, BUA, SI</td>
<td>0.47-0.56</td>
<td>− to +/−</td>
</tr>
<tr>
<td>Weeks et al. (2016)</td>
<td>Calcaneus</td>
<td>4-18</td>
<td>389</td>
<td>Healthy</td>
<td>PCC, LRA, MRA</td>
<td>BUA</td>
<td></td>
<td>−</td>
</tr>
</tbody>
</table>

**Morbidity:** ESRF, End-stage renal failure; Idiopathic SH, Idiopathic subclinical hypothyroidism; BC, breast cancer; CRD, Chronic rheumatic disease; IB, Inflammatory bowel disease; CF, Cystic fibrosis; 21-OHD, 21-hydroxylase deficiency; CKD, Chronic kidney disease.

**Statistical methods:** PCC, Pearson’s correlation coefficient; LCC, linear correlation coefficient; SCC, Spearman’s correlation coefficient; CT, cross tabulation; ROC, receiver operating characteristic analysis; LRA, linear regression analysis; MRA, multiple regression analysis; BAA, Bland-Altman analysis; Cx, Cohen x analysis.

**Authors’ conclusion:** +, good association; +/−, moderate association; −, weak/poor association; − to +/−, moderate to good association; − to +, weak to good association.
References


