ORIGINAL ARTICLE

Significant differences in UK and US female bone density reference ranges

E. Noon · S. Singh · J. Cuzick · T. D. Spector ·
F. M. K. Williams · M. L. Frost · A. Howell ·
M. Harvie · R. Eastell · R. E. Coleman · I. Fogelman ·
G. M. Blake · for the IBIS-II Bone Substudy

Abstract

Summary In the United Kingdom (UK), T- and Z-scores are usually calculated using reference ranges derived from United States (US) populations. In the UK arm of a recent randomised trial (International Breast Cancer Intervention Study II (IBIS-II)), substantially, fewer women than expected were recruited into the osteopenic ($-2.5 <$ T-score < -1.0) and osteoporotic (T-score ≤ -2.5) arms of the study. The comparison with data from two independent studies showed that UK women aged >45 years with a typical body mass index of 28 kg m^{-2} have spine and hip bone mineral density (BMD) 0.6 standard deviation higher than their US counterparts.

Introduction Dual energy X-ray absorptiometry (DXA) is widely used for the diagnosis of osteoporosis and to investigate the effect of pharmacological treatments on BMD. In both routine and research settings, it is important that DXA results are correctly interpreted.

Methods T- and Z-scores for the first 650 UK Caucasian women enrolled in the IBIS-II study were compared with data from two independent studies of unrelated, unselected UK Caucasian women: (1) 2,382 women aged 18 to 79 recruited to the Twins UK Adult Twin Registry; (2) 431 women aged 21 to 84 with no risk factors for osteoporosis recruited at Guy's Hospital. All DXA measurements were performed on Hologic densitometers. Subjects were divided into six age bands, and T- and Z-scores were calculated using the manufacturer's US reference range for the spine and the National Health and Nutrition Examination Survey III reference range for the femoral neck and total hip.

Results The overall mean Z-scores for the IBIS-II, Twin, and Guy's groups were: spine: +0.61, +0.29, +0.33; femoral neck: +0.42, +0.36, +0.45; total hip: +0.65, +0.38, +0.39 (all $p < 0.001$ compared with the expected value of 0). The mean body weight of subjects in the three studies was 74.4, 65.5, and 65.4 kg, respectively. Analysis revealed a highly significant relationship between Z-score and weight at each BMD site with a slope of 0.03 kg^{-1}.

E. Noon · M. L. Frost · I. Fogelman · G. M. Blake
Osteoporosis Research Unit, Division of Imaging Sciences,
King's College London,
London, UK

S. Singh · J. Cuzick
Cancer Research UK, Wolfson Institute of Preventive Medicine,
Queen Mary College London,
London, UK

T. D. Spector · F. M. K. Williams
Department of Twin Research and Genetic Epidemiology,
King's College London,
London, UK

A. Howell
Cancer Research UK Department of Medical Oncology,
Christie Hospital NHS Trust, University of Manchester,
Manchester, UK

A. Howell · M. Harvie
Nightingale and Genesis Prevention Centre,
University Hospital of South Manchester,
Manchester, UK

R. Eastell
Academic Unit of Bone Metabolism, University of Sheffield,
Sheffield, UK

R. E. Coleman
Academic Unit of Clinical Oncology,
Cancer Research Centre Weston Park Hospital,
Sheffield, UK

G. M. Blake (✉)
Osteoporosis Research Unit, 1st Floor,
Tower Wing, Guy's Hospital,
London SE1 9RT, UK
e-mail: glen.blake@kcl.ac.uk

Conclusions In general, US spine and hip reference ranges are not suitable for the calculation of Z-scores in UK women. For some research study designs, the differences may significantly influence the pattern of subject recruitment.

Keywords Body weight · Bone mineral density · Dual energy X-ray absorptiometry · Reference ranges · T-scores · Z-scores

Introduction

Measurements of bone mineral density (BMD) of the spine and hip using dual energy X-ray absorptiometry (DXA) have become the "gold standard" for the diagnosis of osteoporosis [1]. DXA is also used in research studies to investigate the effect of pharmacological treatments on BMD [2–6]. DXA measurements are interpreted using T-scores and Z-scores, where T-scores express the difference in standard deviation units between a subject's BMD and the mean figure in healthy young adults, while Z-scores express the difference from a healthy age-matched reference population [7]. In both clinical and research settings, it is important that BMD results are correctly interpreted, and in the United Kingdom (UK), T- and Z-scores are frequently calculated using reference ranges derived from United States (US) populations. These are the manufacturer's US reference range for the spine [8, 9] and the National Health and Nutrition Examination Survey (NHANES) III reference range for hip sites [10]. Accurate calculation of T- and Z-scores relies upon the gender and ethnic-matched reference population being representative of the local population.

The International Breast Cancer Intervention Study II (IBIS-II) is an international, multicentre, double-blinded, randomised controlled trial of anastrozole vs. placebo in postmenopausal women at increased risk of developing breast cancer. In the IBIS-II Bone Subprotocol Study, a group of subjects enrolled in the main breast cancer prevention trial were invited to take part in an additional study to examine the effectiveness of risedronate at preventing bone loss associated with anastrozole treatment [11]. Women participating in the bone substudy were enrolled into one of three strata: 300 subjects with osteoporosis (T-score $\leq$−2.5) at baseline, who received risedronate 35 mg once weekly; 400 subjects with osteopenia (−2.5 < T-score < −1.0), who were randomised to receive either risedronate or a placebo; and 300 subjects with a normal BMD (T-score $\geq$−1.0), none of whom received bone protection therapy.

During enrolment for the bone substudy, recruitment was slower than expected because of a deficit of subjects enrolled in the osteoporotic and osteopenic groups. Instead of the predicted proportions of 30%:40%:30% for the normal, osteopenic, and osteoporotic groups, the ratio was 57%:33%:10%. To study the reasons for this, T- and Z-scores for the first 650 UK Caucasian women recruited to the IBIS-II Bone Substudy were compared with data from two independent studies of UK women: 2,382 women recruited to the Twins UK Adult Twin Registry [12], and 431 women recruited at Guy's Hospital, London [13]. It was noticed that women enrolled in the IBIS-II Bone Substudy were on average 10 kg heavier than the women in the Twin Registry and Guy's Hospital studies. Since body weight is known to have a significant effect on individual's BMD measurements, the role of body weight and its influence on T- and Z-score levels of the IBIS study women was investigated.

The present study had two aims: (1) to investigate mean T- and Z-scores calculated using US reference ranges in UK Caucasian women enrolled in the IBIS-II Bone Substudy and compare them with T- and Z-scores figures found in two independent studies of healthy UK women; (2) to investigate the effect of body weight on Z-score values and its contribution to the BMD differences between the three study groups.

Subjects and methods

IBIS-II study

Women could be enrolled in the IBIS-II Breast Cancer Prevention Study if they were postmenopausal, had no radiological evidence of breast cancer, and had a twofold relative risk increase of developing breast cancer, which was defined by: (1) relatives having developed breast or ovarian cancer; (2) nulliparous; (3) radiological evidence of benign disease; and other more specific criteria. Subjects with low bone density could only participate if they agreed to take bisphosphonates. The IBIS-II Bone Subprotocol Study enrolled an unselected group of subjects from the main prevention study to assess whether risedronate treatment attenuates the deleterious effects of anastrozole on BMD [14]. The present investigation used the BMD data at enrolment for the first 650 UK Caucasian women scanned on Hologic DXA densitometers taking part in the IBIS-II Bone Subprotocol Study. Women participating in the Bone Subprotocol were a representative sample of the main IBIS-II study.

Twin study

The Twins UK Adult Twin Registry is a volunteer cohort of adult twins aged 18–85 that presently includes approximately 10,000 monozygotic and dizygotic twins, recruited by media campaigns from all over the UK [12]. A database of DXA BMD measurements and demographic

data from the registry was used in this study. Dizygotic twins have DNA that is approximately 50% identical to each of their siblings (twin or not), and monozygotic twins have identical DNA. To rule out any genetic bias, results from just one of each pair of twins having BMD measurements were included in the present investigation, leaving data for 2,382 singleton Caucasian females. Studies have shown that this population is similar to singleton populations derived from age–sex registers such as the Chingford study [15].

Guy's study

This population consisted of 878 Caucasian women recruited from three sources: patients referred by their general practitioner for routine bone density screening by DXA; young hospital personnel; and women from the general population who volunteered to participate in clinical research [13]. Subjects were enrolled for DXA measurements of spine and hip BMD. To obtain a group of healthy women, those with risk factors known to influence bone metabolism or with a history of low-trauma fracture were excluded, leaving 431 subjects [13].

Measurements

BMD measurements of the lumbar spine and proximal femur (nondominant hip) were obtained using Hologic QDR-4500 densitometers (Hologic Inc., Bedford, MA), except for some early data from the Twin study that were obtained using a Hologic QDR-2000. All scans were performed by trained technologists. IBIS-II scans performed at 14 different UK centres were analysed at a central quality control centre at Guy's Hospital. Regions of interest analysed were the total spine (L1–L4), femoral neck, and total hip. Subjects' age, height, and weight were also recorded. All the densitometers were installed and calibrated by the manufacturer's UK representative. A review of daily scans of the Hologic spine phantom on each densitometer showed precision figures that averaged 0.38% (range 0.30% to 0.54%). Spine phantom BMD measurements averaged 100.02% of the nominal values (range 99.61% to 100.30%). A study of long-term in vivo precision gave figures of 1.6%, 2.5%, and 1.6% for spine, femoral neck, and total hip BMD, respectively [16]. A separate in vivo precision study performed for the Twins UK Adult Twin Registry gave precision results of 0.8% at the spine and 1.6% at the femoral neck [17].

Data analysis

The data were divided into six age bands: <25, 25–34, 35–44, 45–54, 55–64, and ≥65 years. Weight, height, and body mass index (BMI = weight (kg)/height (m)^2) data for the three study groups were compared with recently published figures for the UK female population [18] and anthropometric data for non-Hispanic white women from the US NHANES III study [19]. Z-scores were calculated for all subjects by the difference between the subject's and the age-matched mean BMD, expressed in standard deviation (SD) units. T-scores were calculated with respect to the young adult reference values. Spine T- and Z-scores were calculated using the manufacturer's US spine reference range [8, 9], and hip scores were calculated using the NHANES III range [10]. DXA manufacturers have adopted the NHANES levels as their reference data for calculating T- and Z-scores at the proximal femur [20].

The mean and SD of the T- and Z-scores values were calculated for each age group in each study population to demonstrate gross differences from the reference populations for the spine, femoral neck, and total hip regions of interest. If the reference range correctly describes the study population, the mean Z-score will be Z=0 and SD=1.0. T- and Z-scores values are related by the following equation [21]:

$$\text{T-score} = \text{age-specific mean T-score} + \text{Z-score} \qquad (1)$$

In Eq. 1, the age-specific mean T-score is the curve of mean T-score as a function of age obtained by expressing the BMD reference range in T-score units [22]. Student's t tests were performed to examine whether the mean Z-scores were significantly different from 0 [23]. The Z-score SDs were tested against the reference population SD of 1.0 using the F test [23]. Statistical significance was based on a p value of <0.05.

To examine the effect of body weight on BMD, each subject's Z-score was plotted against their weight and a linear regression line fitted to the points [23]. The plot was performed for each study group at each BMD site. The weight coefficient—the gradient of this regression line—was calculated for each plot together with the intercept on the weight axis.

Results

Demographic data for the three study groups including figures for weight, height, and BMI broken down by age are shown in Table 1. The absence of subjects aged <45 years in the IBIS group reflects the entry requirement to be postmenopausal. Ages in the IBIS group ranged from 45 to 71 years, for the Twin study from 18 to 79 years, and for the Guy's study from 21 to 84 years. The number of subjects in each age group varied from 11 to 742 (Table 1).

Table 1 Demographic data for the three study groups including figures for weight, height, and BMI broken down by age

	IBIS-II Bone Substudy				Twin study				Guy's study			
Age	n	W	H	BMI	n	W	H	BMI	n	W	H	BMI
<25					142	63.1	1.65	23.1	11	69.8	1.71	23.8
25–34					315	64.8	1.64	24.1	45	64.6	1.66	23.5
35–44					496	65.4	1.63	24.5	63	67.2	1.64	25.0
45–54	184	74.2	1.63	28.0	742	65.9	1.63	25.0	160	64.6	1.64	24.1
55–64	372	74.3	1.62	28.2	534	66.7	1.61	25.6	94	66.8	1.62	25.6
≥65	94	75.3	1.61	29.1	153	64.1	1.59	25.4	58	64.1	1.58	25.6
Tot/Av	650	74.4	1.62	28.3	2,382	65.5	1.63	24.8	431	65.4	1.63	24.7

n number in group, *W* weight (kilogram), *H* height (metre), *BMI* body mass index (kilogram per square metre), *Tot/Av* total *n*, average W, H, or BMI

The Twin study and Guy's study have similar overall mean weights of 65.5 and 65.4 kg, respectively, while the IBIS study was higher at 74.4 kg. Except for the <25-year group in the Guy's study (*n*=11), mean weights in the Twin and Guy's study groups ranged between 63.1 and 67.2 kg. Mean weights in the IBIS study group were between 74.2 and 75.3 kg, consistently higher than the other two studies.

Table 2 summarises recently published weight, height, and BMI data for UK women [18] together with similar data for non-Hispanic white females from the US NHANES III study [19]. For the IBIS study, the mean BMI of women was 28.3 kg/m^2. For comparison, the mean BMI for women aged >45 years was 27.9 kg/m^2 for UK women, 26.8 kg/m^2 for US women in the NHANES III study, 25.3 kg/m^2 for the Twin study, and 24.8 kg/m^2 for the Guy's study. For women <45 years in the Twin and Guy's studies, the mean BMI figures were closer to the population values for UK and US women. Mean heights for women >45 years in all three studies were similar to the population figures for UK and US women (IBIS study: 162.2 cm; Twin study: 161.7 cm; Guy's study: 162.0 cm; UK women: 160.6 cm; US women 161.1 cm). Mean weights for women >45 years

reflected the differences found in BMI (IBIS study: 74.4 kg; Twin study: 66.0 kg; Guy's study: 65.2 kg; UK women: 71.5 kg; US women 67.8 kg).

Mean T- and Z-score vs. age

Spine

Mean T-scores for each age group in all three studies are shown plotted in Fig. 1a together with the manufacturer's spine reference range. Although for the 25–34-year age group there is a good agreement with the reference data, at other ages, there is a trend for increasing positive offsets from the reference curve with increasing age. The mean overall Z-scores expressing the divergence from the reference range Eq. 1 were +0.61, +0.29, and +0.33 for the IBIS-II, Twin, and Guy's studies, respectively. For women >45 years in the Twin and Guy's studies, the mean Z-scores were +0.44 and +0.35, respectively. The mean Z-score of −0.28 for the <25-year group in the Twin study was the only statistically significant negative figure in the entire study. For the 35–44-year and older age groups, the mean Z-scores were statistically significantly greater than 0

Table 2 Weight, height, and BMI data

	UK—NHS [18]				US—NHANES III [19]			
Age	n	W	H	BMI	Age	n	W	H	BMI
16–24	294	65.0	1.63	24.4	20–29	452	63.7	1.64	23.8
25–34	453	68.1	1.63	25.7	30–39	579	68.9	1.65	25.5
35–44	649	71.3	1.63	26.8	40–49	477	70.5	1.64	26.4
45–54	527	71.8	1.62	27.4	50–59	484	74.3	1.63	28.2
55–64	538	72.9	1.61	28.2	60–69	501	70.7	1.61	27.3
65–74	393	71.2	1.59	28.0	70–79	643	66.8	1.58	26.7
≥75	281	64.9	1.55	26.9	≥80	621	60.4	1.55	25.1
Total/Av	3,135	69.7	1.61	26.8	Total/Av	3,757	68.5	1.62	26.0

NHANES III data are for non-Hispanic white females

n number in group, *W* weight (kilogram), *H* height (metre), *BMI* body mass index (kilogram per square metre), *Tot/Av* total *n*, average W, H, or BMI
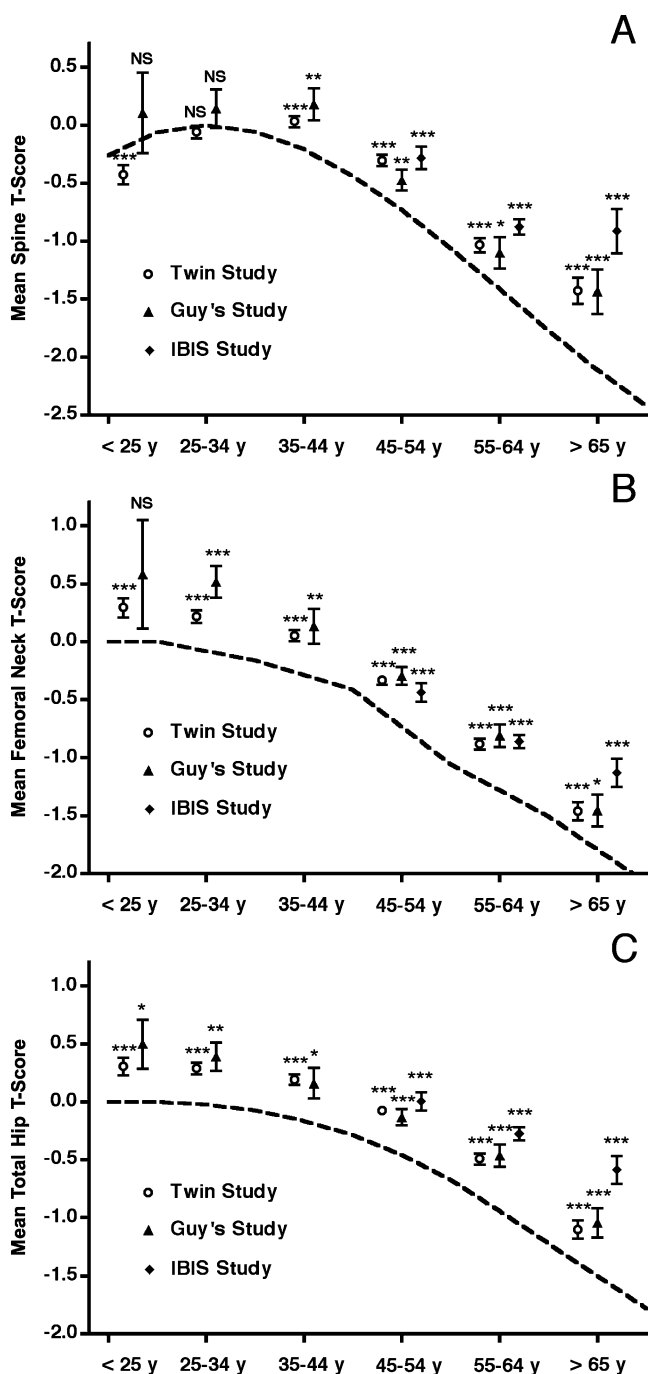
Fig. 1 Mean T-score vs. age for **a** lumbar spine, **b** femoral neck, and **c** total hip BMD for subjects in the three study groups. *Error bars* show ± 1 SEM. *Dashed curves* show the age-specific mean T-score (see Eq. 1) calculated from the BMD reference ranges. Spine reference range: Hologic manufacturer's [8, 9]; femoral neck and total hip: NHANES III [10]. *Asterisks* show the statistical significance of the difference of the mean Z-score from zero (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$)

in all three studies. For the young normal group (spine: 25 to 34 years), the mean T-score (standard error of the mean (SEM)) for the pooled data from the Twin and Guy's studies ($n = 360$) was −0.05 (0.05).

Femoral neck

Mean T-scores for each age group and each study are plotted in Fig. 1b. At the femoral neck site, the mean Z-scores were all statistically significantly greater than 0 except for the <25-year age group in the Guy's study. In contrast to the trend seen in the spine, the hip sites show a similar offset from the NHANES III reference curve across all age groups. The mean overall Z-scores were +0.42, +0.36, and +0.45 for the IBIS-II, Twin, and Guy's studies, respectively. For women >45 years in the Twin and Guy's studies, the mean Z-scores were +0.39 and +0.43, respectively. For the young normal group (hip: 20 to 29 years), the mean T-score (SEM) for the pooled data from the Twin and Guy's studies ($n = 276$) was +0.32 (0.06).
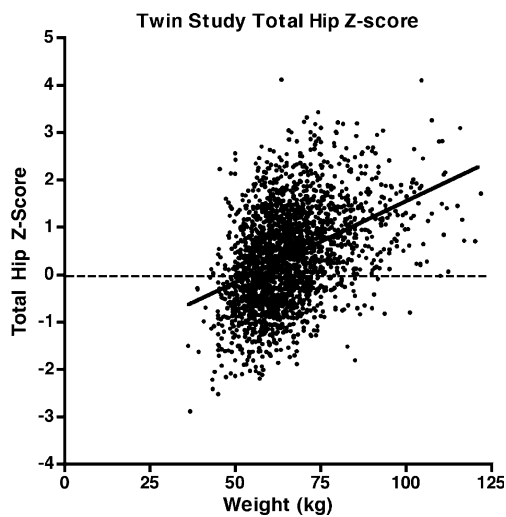
Total hip

Mean T-scores for each age group and each study are plotted in Fig. 1c and are systematically higher than the NHANES III reference curve. The mean Z-score results for each age group were all statistically significantly greater than 0. The mean overall Z-scores for the total hip site were +0.65, +0.38, and +0.39 for the IBIS-II, Twin, and Guy's studies, respectively. For women >45 years in the Twin and Guy's studies, the mean Z-scores were +0.41 and +0.40, respectively. For the young normal group, the mean T-score (SEM) for the pooled data from the Twin and Guy's studies was +0.34 (0.06).

In the IBIS study, Z-score data hip SDs were within ±0.18 of 1.0, but the spine showed consistently higher SDs rising with age from 1.29 to 1.84. In the Twin study, the standard deviation of Z-scores was within ±0.12 SDs from 1.0, except in the older age groups at the spine where the SD increased from 1.27 at 45–54 years to 1.40 in the oldest subjects. A small sample size increases the statistical error, so results from the <25-year age group in the Guy's study ($n = 11$) are not significant. Otherwise, the Z-score SD was within ±0.2 SD from 1.0, except for the older spine groups where it increased from 1.31 in the 55–64-year group to 1.47 in the ≥65-year group.

Z-score vs. weight

Subjects' Z-scores were plotted against their weight for each study and each region of interest, and a linear regression line was fitted to these points. Figure 2 is a representative example of these plots. Gradients of the regression lines for the Z-score vs. weight plots were calculated and are shown in Fig. 3a. All the gradients were statistically significantly different from 0 ($p < 0.001$) and within the statistical errors approximated to the same figure (chi-square=5.8; df=8; $p = 0.67$). The weighted mean gradient was 0.034 kg^{-1}.

Fig. 2 Scatter plot of total hip Z-score vs. body weight for subjects in the Twin study. The *straight line* shows the regression line fitted to the data
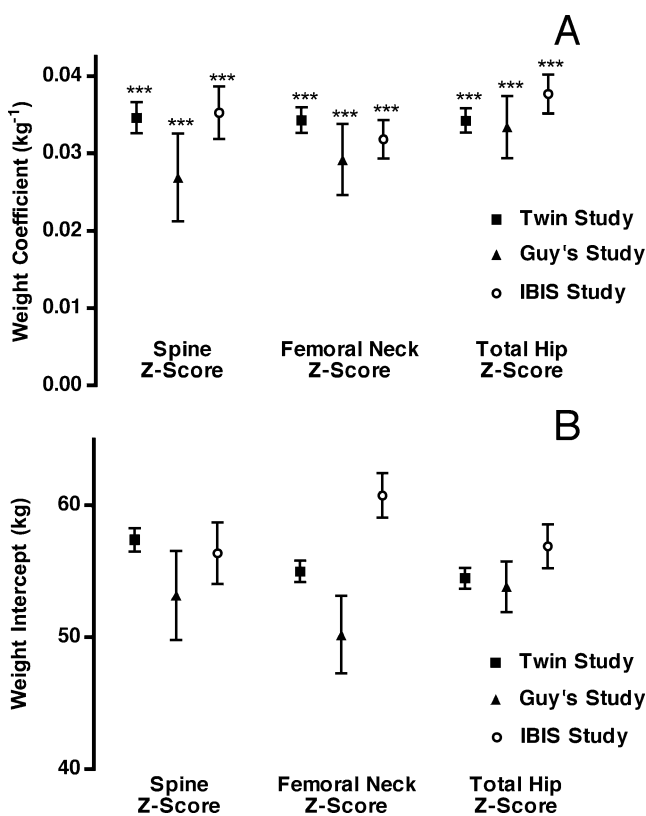
Fig. 3 **a** Plot of the slope of the Z-score vs. body weight regression line for spine, femoral neck, and total hip Z-score for subjects in the three study groups (***$p < 0.001$). An example of a Z-score vs. weight regression line is shown as the *solid line* in Fig. 2. **b** Plot of the intercept on the weight axis of the Z-score vs. body weight regression line for spine, femoral neck, and total hip Z-score for subjects in the three study groups. The weight intercept is the intercept of the regression line (*solid line* in Fig. 2) on the *dashed line (Z-score= 0) in* Fig. 2

Intercept points on the weight axis were also calculated from the regression lines in the Z-score vs. weight plots (Fig. 3b). These equate to the weight in kilogram that gives a mean Z-score of 0. The largest of the intercepts was 60.8 kg in the femoral neck from the IBIS study, and the smallest was 50.2 kg in the femoral neck from the Guy's study. Overall, the mean intercept was 55.3 kg.

Discussion

Subjects in the Twin and Guy's studies were nonselected, included a wide range of ages, and were healthy UK Caucasian women. Using subjects from two independent studies makes errors due to DXA calibration, operator technique, or subject selection bias less likely. The issue of whether women with a diagnosis of breast cancer have higher than average BMD is controversial [24, 25]. However, an increased risk of breast cancer of itself is not known to affect BMD. Although all three groups might be considered representative of the UK female population, a weight difference of almost 10 kg was found between the IBIS subjects and the Twin and Guy's groups that would be expected to influence the BMD findings. Comparison with UK and NHANES III anthropometric data shows that the mean BMI of 28.3 kg/m^2 for the IBIS study women is similar to the figure for UK and US women aged >45 years, while the Twin and Guy's study women >45 years had a significantly lower BMI than the UK average. In contrast, for women <45 years, the BMI data for the Twin and Guy's study women were in better agreement with the UK and US female reference data. Average heights for the women >45 years in the three study groups of around 1.62 m were similar to the UK and US norms, and so, the differences in BMI largely reflect the differences in body weight.

In the UK, T- and Z-scores are usually calculated using reference ranges derived from US populations. Historically, there was poor agreement between hip T-scores calculated on GE-Lunar and Hologic densitometers [26]. The NHANES III study included over 14,000 randomly selected individual's representative of the whole US population, making subject selection bias in this reference range unlikely [10]. This extensive sampling of the US population led the principal DXA manufacturers to adopt the NHANES III BMD data for the calculation of T- and Z-scores [20]. In contrast, the Hologic spine reference range was generated from volunteers scanned at some of the first DXA centres in the US and includes measurements of 605 women [8, 9]. The mean body weight for the women aged >45 years was 63.7 kg [9]. Despite the small sample size and coverage compared with the NHANES III

study, there is good agreement between Hologic and GE-Lunar spine T-scores calculated using the manufacturers' reference ranges [26].

Reference ranges used to calculate Z-scores take into account the bone loss associated with ageing, and at each age the BMD data are assumed to follow a normal distribution with the same population SD. Therefore, BMD measurements from a large population that match the reference data should have a mean Z-score of 0 and an SD of 1.0 at all ages. In Fig. 1, the mean Z-score values are shown as the offset of the mean T-score values from the dashed curves showing the reference levels (Eq. 1). At both the femoral neck and total hip sites, the mean Z-scores were statistically significantly different from 0 for all ages and all studies, except the <25-year group in the Guy's study ($n=$ 11), with overall mean values of 0.36–0.65 that remained relatively constant across the full age range (Fig. 1b, c). At the spine, the majority of mean Z-scores were statistically significant from 0, though three groups were not significant, and one gave a statistically significant negative figure. In contrast to the hip sites, for the spine, there was a trend for the mean Z-score to increase with age. When the Hologic spine reference range was constructed, spinal radiographs were taken to exclude subjects with degenerative disease [8]. Therefore, some of the deviation seen in Fig. 1a might be explained by the increased prevalence of degenerative disease in older subjects. In the Twin and Guy's studies, women >45 years had a lower overall mean Z-score at the spine and total hip sites than the IBIS-II subjects (spine: 0.35–0.44 vs. 0.61; total hip: 0.40–0.41 vs. 0.65). The demographic data suggests this might be due to their lower body weight compared with the IBIS-II study. For the young adult subjects in the Twin and Guy's studies, the pooled mean T-scores were in close agreement with the expected value of 0 for spine BMD but, for the femoral neck and total hip, were statistically significantly different at +0.32 and +0.34, respectively.

For the hip regions, the Z-score SDs were constant with age and were not statistically significantly different from 1.0. However, in subjects aged >45 years, spine SDs differed significantly from 1.0 with the SD figure increasing with age. There were no gross differences in the SD figures between studies.

The slopes of the regression lines fitted to the scatter plots of Z-score vs. weight showed little variation between the three studies or between the different BMD sites and were consistent with an overall mean figure of 0.034 kg^{-1} (Fig. 3a). This finding confirms previous evidence that body weight significantly influences BMD and therefore, Z-score [27]. The mean weight intercept was 55 kg (Fig. 3b), which reflects the body weight at which the US reference ranges would give a mean Z-score figure of 0 in UK women.

Several previous studies have reported similar Z-score discrepancies in UK women. Petley et al. investigated BMD measurements from GE-Lunar densitometers in 700 women from Southampton [28]. They found higher BMD results by about 0.5 SDs in their study population compared with the manufacturer's US reference ranges for spine, femoral neck, and total body BMD (all age points $p<0.05$ or less). Similar results were found for a large multicentre database of UK women [29] and also for the Chingford study [15]. The latter included approximately 900 women aged 45–65 years recruited from general practice who were found to have a mean spine Z-score of +0.35 and a mean femoral neck Z-score of +0.16 [15], figures that are both statistically significantly different from 0.

Studies of other Caucasian populations show geographical variation in spine and hip BMD [9, 30–32]. The strongest evidence is from the DXA bone density measurements performed during the European Vertebral Osteoporosis Study (EVOS) [30]. A comparison of BMD measurements in 2,228 men and 2,581 women aged 50–80 years from 16 centres in 12 European countries showed highly significant differences between centres after adjustment for age, height, and weight. The between centre SD was approximately 25% of the population SD, and the difference in mean BMD between the highest and lowest centres was about 1 SD. The authors concluded that the study demonstrated clear differences between European populations in BMD at the spine and hip. Kaptoge et al. reported similar geographical BMD variations among young European men and women [31]. In a study of 1,163 men and 329 women aged 19–29 years from 17 centres in 12 countries, they reported significant variation in adjusted mean spine and hip BMD levels arising from national differences. Although mean European values for hip BMD differed from US figures, SDs were indistinguishable. Flicker et al. reported BMD differences between Australian and North American white female populations based on measurements performed on Hologic instruments [9]. Measurements of 411 Australian women aged 15–74 years gave a mean spine Z-score of +0.63 and a mean femoral neck Z-score of +0.37 when interpreted using the manufacturer's North American spine reference range and the NHANES III hip reference range, respectively. A recent study of Australian men from the Geelong Osteoporosis Study showed femoral neck BMD levels similar to those obtained in the US non-Hispanic white males participating in the NHANES III study. However, spine BMD levels differed from figures from France, Canada, and Spain even after exclusion of men with vertebral abnormalities on spinal radiographs [32].

This study has emphasised body weight as an important and often neglected factor influencing subjects'

spine and hip BMD results. The two principal DXA manufacturers calculate Z-scores differently, GE-Lunar making an allowance for body weight but not Hologic [33]. The long-term trend towards increasing obesity raises the question of what effect this might have on BMD measurements. Data for UK women show that over the 12-year period from 1993 to 2004, mean body weight over all age groups increased from 66.6 to 69.7 kg, height from 161.1 to 161.4 cm, and BMI from 25.7 to 26.8 kg/m^2 [18]. There is a steady trend for weight to increase by 0.3 kg/year, corresponding to a predicted Z-score increase of 0.1 per decade.

The results of the present study make it clear that the reason for the slow recruitment of UK women to the osteoporotic and osteopenic arms of the IBIS-II Bone Substudy was the unexpectedly high Z-scores of UK women in the >45-year age group. With an average body weight of 74 kg, the IBIS-II subjects were the typical weight of UK women in their age group and similar to the non-Hispanic US white women in the NHANES III study. A difference in body weight is, therefore, not the explanation for the Z-score differences between UK and US women. In contrast, the mean body weight of 65 kg found in the Twin and Guy's study women >45 years is not typical of older UK women [18]. With the weight coefficient of 0.03 kg^{-1} found from the Z-score vs. weight scatter plots, and the women in the Twin and Guy's studies had the higher body weight found in the IBIS-II and NHANES III studies, the Z-score differences between UK and US women would have increased by about 0.2–0.3 SD from around 0.4 SD to around 0.6 SD. However, higher body weight did not explain all the differences between the IBIS-II study Z-scores and the other two studies. At all three BMD sites, there was a trend for mean Z-score to increase with age in the IBIS-II women despite the fact that their mean body weight did not change.

Even at the lower than average body weight of 65 kg found in the Twin and Guy's studies, mean spine, femoral neck, and total hip Z-scores in UK women calculated using US reference ranges lie in the range 0.3 to 0.4. Only for groups of women with a mean body weight as low as 55 kg would the US reference ranges give appropriate Z-score values in UK women. With the majority of mean Z-score results significantly greater than 0 in all three study populations, we conclude that, in general, US reference ranges are not reliable for calculating Z-scores values for UK Caucasian women.

In the IBIS-II Bone Substudy and other similar research studies involving the selection of subjects based on a BMD measurement, the assignment of individuals to different study groups is often based on the interpretation of the BMD findings in terms of T-scores. It is

better that a common reference range such as NHANES III is used for such selection procedures, since the use of different reference ranges for studies performed in different countries can only add to the confusion. Nevertheless, whether the adopted reference range is actually a good description of the population screened for the study can have a significant effect on the enrolment of subjects. In the IBIS-II Bone Study, the unexpectedly high overall mean Z-scores of 0.42–0.65 were sufficient to delay significantly recruitment into the osteoporotic and osteopenic groups.

The choice of reference ranges for calculating T-scores for the diagnosis of osteoporosis has always been a cause of controversy. It is not the intention of the present article to suggest that UK DXA centres should not continue to use US reference ranges for the diagnosis of osteoporosis based on T-scores. In this application, it is more important that an accepted reference range such as NHANES III is adopted as widely as possible rather than cause confusion by using inconsistent local reference ranges for calculating T-scores. For example, the use of NHANES III femoral neck T-scores is a requirement when T-scores are used in the World Health Organisation FRAX tool for evaluating patients' 10-year risk of osteoporotic fracture [34]. The availability of the FRAX tool renders the previous controversies over which reference to use less relevant, since DXA measurements can be entered as raw BMD results and the relationship between BMD and fracture risk is derived from population and ethnicity-dependent epidemiological data rather than comparison with reference ranges. Z-scores are used to compare patients with their peers in terms of healthy controls matched for age, gender, and ethnic origin and for GE-Lunar DXA systems, body weight too. To this extent, clinicians should be aware that the use of US reference ranges to calculate spine and hip Z-scores for UK women with body weights typical of the population may overestimate the true Z-score by 0.6 SD. Pan-European studies such as EVOS have found significant BMD differences between countries, and so, expressed in terms of Z-score offsets from US populations, the exact correction will differ from country to country.

This study has several limitations. We used Z-scores calculated from Hologic densitometers only since insufficient data were available to evaluate Z-scores on GE-Lunar systems. Also, this investigation was limited to data from Caucasian women alone, and there were no data for male subjects.

In conclusion, the US spine and hip reference ranges are generally not suitable for the calculation of Z-scores in UK women, and for some research study designs, these differences may significantly influence the pattern of subject recruitment. The results confirmed that body weight has an

important influence on spine and hip BMD measurements, and that for studies enrolling groups of subjects with atypical BMI, this may be an additional factor affecting recruitment.

Conflicts of interest None.

References

1. Kanis JA, McCloskey EV, Johansson H, Oden A (2008) Review: a reference standard for the description of osteoporosis. Bone 42:467–475

2. Black DM, Cummings SR, Karpf DB, Cauley JA, Thompson DE, Nevitt MC, Bauer DC, Genant HK, Haskell WL, Marcus R, Ott SM, Torner JC, Quandt SA, Reiss TF, Ensrud KE (1996) Randomised trial of the effect of alendronate on risk of fracture in women with existing vertebral fractures. Lancet 348:1535–1541

3. Reid IR, Brown JP, Burckhardt P, Horowitz Z, Richardson P, Trechsel U, Widmer A, Devogelaer JP, Kaufman JM, Jaeger P, Body JJ, Brandi ML, Broell J, Di Micco R, Genazzani AR, Felsenberg D, Happ J, Hooper MJ, Ittner J, Leb G, Mallmin H, Murray T, Ortolani S, Rubinacci A, Saaf M, Samsioe G, Verbruggen L, Meunier PJ (2002) Intravenous zoledronic acid in postmenopausal women with low bone density. N Engl J Med 346:653–661

4. Meunier PJ, Roux C, Seeman E, Ortolani S, Badurski JE, Spector TD, Cannata J, Balogh A, Lemmel EM, Pors-Nielsen S, Rizzoli R, Genant HK, Reginster JY (2004) The effects of strontium ranelate on the risk of vertebral fracture in women with postmenopausal osteoporosis. N Engl J Med 350:459–468

5. Greenspan SL, Bone HG, Ettinger MP, Hanley DA, Lindsay R, Zanchetta JR, Blosch CM, Mathisen AL, Morris SA, Marriott TB (2007) Effect of recombinant human parathyroid hormone (1-84) on vertebral fracture and bone mineral density in postmenopausal women with osteoporosis: a randomised trial. Ann Intern Med 146:326–339

6. Black DM, Delmas PD, Eastell R, Reid IR, Boonen S, Cauley JA, Cosman F, Lakatos P, Leung PC, Man Z, Mautalen C, Mesenbrink P, Hu H, Caminis J, Tong K, Rosario-Jansen T, Krasnow J, Hue TF, Sellmeyer D, Eriksen EF, Cummings SR (2007) Once-yearly zoledronic acid for treatment of postmenopausal osteoporosis. N Engl J Med 356:1809–1822

7. Blake GM, Fogelman I (2007) The role of DXA bone density scans in the diagnosis and treatment of osteoporosis. Postgrad Med J 83:509–517

8. Kelly TL (1990) Bone mineral density reference databases for American men and women. J Bone Miner Res 5:249

9. Flicker L, Green R, Kaymakci B, Lichtenstein M, Buirski G, Wark JD (1995) Do Australian women have greater spinal bone density than North American women? Osteoporos Int 5:63–65

10. Looker AC, Wahner HW, Dunn WL et al (1998) Updated data on proximal femur bone mineral levels of US adults. Osteoporos Int 8:468–489

11. Singh S, Cuzick J, Edwards R, Blake G, Truscott J, Coleman J, Eastell R, Howell A (2007) Effect of anastrozole on bone mineral density after one year of treatment: results from the bone sub-study of the International Breast Cancer Intervention Study (IBIS-II). Abstracts from the 30th Annual San Antonio Breast Cancer Symposium Dec 13-16 2007. Available online at http://www.sabcs.org/EnduringMaterials/ Accessed 23 October 2009.

12. Spector TD, Williams FMK (2006) The UK adult twin registry (TwinsUK). Twin Res Hum Genet 9:899–906

13. Frost ML, Blake GM, Fogelman I (2000) Can the WHO criteria for diagnosing osteoporosis be applied to calcaneal quantitative ultrasound? Osteoporos Int 11:321–330

14. Eastell R, Adams JE, Coleman RE, Howell A, Hannon RA, Cuzick J, Mackey JR, Beckmann MW, Clack G (2008) Effect of anastrozole on bone mineral density: 5-year results from the anastrozole, tamoxifen, alone or in combination trial 18233230. J Clin Oncol 26:1051–1057

15. Andrew T, Hart DJ, Snieder H, de Lange M, Spector TD, MacGregor AJ (2001) Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. Twin Res 4:464–477

16. Patel R, Blake GM, Rymer J, Fogelman I (2000) Long-term precision of DXA scanning assessed over seven years in forty postmenopausal women. Osteoporos Int 11:68–75

17. Zhai G, Andrew T, Kato BS, Blake GM, Spector TD (2009) Genetic and environmental determinants on bone loss in post-menopausal Caucasian women: a 14-year longitudinal twin study. Osteporosis Int 20:949–953

18. Health Survey for England 2004 – updating of trend tables to include 2004 data. The Health and Social Care Information Centre, 2005. Available online at: http://www.ic.nhs.uk/statistics-and-data-collections/health-and-lifestyles-related-surveys/health-survey-for-england/health-survey-for-england-2004-updating-of-trend-tables-to-include-2004-data (accessed 23 October 2009)

19. McDowell MA, Fryar CD, Ogden CL (2009) Anthropometric reference data for children and adults: United States, 1988-1994. National Center for Health Statistics. Vital Health Stat, Series 11, No. 249

20. Hanson J (1997) Standardization of femur BMD. J Bone Miner Res 12:1316–1317

21. Blake GM, Chinn DJ, Steel SA, Patel R, Panayiotou E, Thorpe J, Fordham JN (2005) A list of device specific thresholds for the clinical interpretation of peripheral X-ray absorptiometry examinations. Osteoporos Int 16:2149–2156

22. Faulkner KG, Von Stetton E, Miller P (1999) Discordance in patient classification using T-scores. J Clin Densitom 2:343–350

23. Altman DG (1991) Practical statistics for medical research. Chapman Hall, London

24. Cauley JA, Lucas FL, Kuller LH, Vogt MT, Browner WS, Cummings SR (1996) Bone mineral density and risk of breast cancer in older women: the study of osteoporotic fractures. JAMA 276:1404–1408

25. Stewart A, Kumar V, Torgerson DJ, Fraser WD, Gilbert FJ, Reid DM (2005) Axial BMD, change in BMD and bone turnover do not predict breast cancer incidence in early postmenopausal women. Osteoporos Int 16:1627–1632

26. Faulkner KG, Roberts LA, McClung MR (1996) Discrepancies in normative data between Lunar and Hologic DXA systems. Osteoporos Int 6:432–436

27. De Laet C, Kanis JA, Oden A et al (2005) Body mass index as a predictor of fracture risk: a meta-analysis. Osteoporos Int 16:1330–1338

28. Petley GW, Cotton AM, Murrills AJ, Taylor PA, Cooper C, Cawley MI, Wilkin TJ (1996) Reference ranges of bone mineral density for women in southern England: the impact of local data on the diagnosis of osteoporosis. Br J Radiol 69:655–660

29. Holt G, Khaw KT, Reid DM, Compston JE, Bhalla A, Woolf AD, Crabtree NJ, Dalzell N, Wardley-Smith B, Lunt M, Reeve J (2002) Prevalence of osteoporotic bone mineral density at the hip in Britain differs substantially from the US over 50 years of age: implications for clinical densitometry. Br J Radiol 75:736–742

30. Lunt M, Felsenberg D, Adams J, Benevolenskaya L, Cannata J, Dequeker J, Dodenhof C, Falch JA, Johnell O, Khaw KT,

Masaryk P, Pols H, Poor G, Reid D, Scheidt-Nave C, Weber K, Silman AJ, Reeve J (1997) Population-based geographic variations in DXA bone density in Europe: the EVOS Study. Osteoporos Int 7:175–189

31. Kaptoge S, da Silva JA, Brixen K, Reid DM, Kröger H, Nielsen TL, Andersen M, Hagen C, Lorenc R, Boonen S, de Vernejoul MC, Stepan JJ, Adams J, Kaufman JM, Reeve J (2008) Geographical variation in DXA bone mineral density in young European men and women. Results from the Network in Europe on Male Osteoporosis (NEMO) study. Bone 43:332–339

32. Henry MJ, Pasco JA, Korn S, Gibson JE, Kotowicz MA, Nicholson GC (2009) Bone mineral density reference ranges for Australian men: Geelong Osteoporosis Study. Osteoporos Int. Published online 26 Aug 2009

33. Carey JJ, Delaney MF, Love TE, Richmond BJ, Cromer BA, Miller PD, Manilla-McIntosh M, Lewis SA, Thomas CL, Licata AA (2007) DXA-generated Z-scores and T-scores may differ substantially and significantly in young adults. J Clinical Densitom 10:351–358

34. FRAX-WHO Fracture Risk Assessment Tool. Available online at http://www.shef.ac.uk/FRAX. Accessed 23 October 2009