The Choice of Normative Pediatric Reference Database Changes Spine Bone Mineral Density Z-Scores But Not the Relationship Between Bone Mineral Density and Prevalent Vertebral Fractures

Jinhui Ma, Kerry Siminoski, Nathalie Alos, Jacqueline Halton, Josephine Ho, Brian Lentle, MaryAnn Matzinger, Nazih Shenouda, Stephanie Atkinson, Ronald Barr, David A. Cabral, Robert Couch, Elizabeth A. Cummings, Conrad V. Fernandez, Ronald M. Grant, Celia Rodd, Anne Marie Sbrocchi, Maya Scharke, Frank Rauch, and Leanne M. Ward, on behalf of the Canadian STOPP Consortium

Children’s Hospital of Eastern Ontario Research Institute (J.M., M.S.), Ottawa, ON; University of Alberta (K.S., R.C.), Edmonton, AB; Université de Montréal (N.A.), Montréal, QC; University of Ottawa (J.H., M.M., N.S., L.M.W.), Ottawa, ON; University of Calgary (J.H.), Calgary, AB; University of British Columbia (B.L., D.A.C.), Vancouver, BC; McMaster University (S.A., R.B.), Hamilton, ON; Dalhousie University (E.A.C., C.V.F.), Halifax, NS; University of Toronto (R.M.G.), Toronto, ON; University of Manitoba (C.R.), Winnipeg, MB; McGill University (A.M.S., F.R.), Montréal, QC

Objectives: Our objectives were to assess the magnitude of the disparity in lumbar spine bone mineral density (LSBMD) Z-scores generated by different reference databases and to evaluate whether the relationship between LSBMD Z-scores and vertebral fractures (VF) varies by choice of database.

Patients and Design: Children with leukemia underwent LSBMD by cross-calibrated dual-energy x-ray absorptiometry, with Z-scores generated according to Hologic and Lunar databases. VF were assessed by the Genant method on spine radiographs. Logistic regression was used to assess the association between fractures and LSBMD Z-scores. Net reclassification improvement and area under the receiver operating characteristic curve were calculated to assess the predictive accuracy of LSBMD Z-scores for VF.

Results: For the 186 children from 0 to 18 years of age, 6 different age ranges were studied. The Z-scores generated for the 0 to 18 group were highly correlated ($r = 0.90$), but the proportion of children with LSBMD Z-scores $\leq -2.0$ among those with VF varied substantially (from 38–66%). Odds ratios (OR) for the association between LSBMD Z-score and VF were similar regardless of database (OR = 1.92, 95% confidence interval 1.44, 2.56 to OR = 2.70, 95% confidence interval 1.70, 4.28). Area under the receiver operating characteristic curve and net reclassification improvement ranged from 0.71 to 0.75 and $-0.15$ to 0.07, respectively.

Conclusions: Although the use of a LSBMD Z-score threshold as part of the definition of osteoporosis in a child with VF does not appear valid, the study of relationships between BMD and VF is valid regardless of the BMD database that is used. (J Clin Endocrinol Metab 100: 1018–1027, 2015)
Children with a variety of chronic conditions are at risk for bone fragility and reductions in bone mineral density (BMD) due to osteoporosis, resulting either from the underlying conditions (such as leukemia, inflammatory conditions, and disorders that impede ambulation) or their treatment (for example, glucocorticoid therapy) (1–5). Dual-energy x-ray absorptiometry (DXA) is the most widely used technique for the evaluation of BMD in children at risk for fractures, given the rapidity of scan acquisition, minimal radiation dose, the excellent precision and accuracy of the measurement, and broad availability. Interest in the clinical utility of DXA-based BMD measurement has heightened in children over recent years, because a growing number of pediatric studies have shown a clear relationship between low BMD and the risk of vertebral (6–8) and nonvertebral (9) fractures. As a result, the optimal use of DXA-based BMD to identify which children are in need of bone health monitoring is an ongoing point of focus for pediatric bone healthcare providers, with the ultimate goals being to predict individuals at risk of overt bone fragility and to intervene to prevent fractures.

One of the challenges facing clinicians in the use of DXA for assessing BMD in children is choosing the normative database that will be used to convert raw BMD scores to gender- and age-specific Z-scores. Over a dozen published normative databases are available for children on different DXA machines, with Hologic-based studies confirming that Z-scores vary significantly in children depending upon the normative reference database that is used to generate the Z-scores (10, 11). This issue is particularly relevant given that the International Society for Clinical Densitometry (ISCD) pediatric osteoporosis definition includes a BMD threshold of −2.0 SDs or lower along with a clinically significant fracture history; recently, the ISCD proposed that the exception to this is the presence of a low-trauma vertebral fracture (VF), in which case the BMD threshold criteria do not apply (12). The use of the BMD Z-score to define osteoporosis in children raises the importance of understanding the clinical significance of the BMD Z-score variability generated by different normative databases and the relationship between Z-scores and the risk of fragility fractures.

We have previously shown in children with acute lymphoblastic leukemia (ALL) that 16% had prevalent VF around the time of diagnosis, and that every 1.0 SD reduction in lumbar spine BMD (LSBMD) Z-score was associated with 80% increased odds of VF at that time point (6). We further showed that 16% of children had incident (ie, new) VF at 12 months after chemotherapy initiation and that LSBMD Z-score predicted incident VF, with 40% increased likelihood of a new VF at 12 months for every 1 SD reduction in LSBMD Z-score at diagnosis (7).

It is important to note that these observations are based on the use of a specific LSBMD reference database to generate the Z-scores (6, 7). Because it has been previously shown (using Hologic instruments) that LSBMD Z-scores vary considerably depending on the normative database that is used (10, 11), we sought to more fully understand the clinical and research implications of the disparities in Z-scores generated by different BMD reference databases, through study of the relationship between LSBMD Z-scores and a key clinical endpoint in the pediatric chronic illness population, ie, VF. Specifically, our goals were to 1) assess the magnitude of the disparities in LSBMD Z-scores generated by different published normative references using both Hologic and Lunar instruments and 2) evaluate whether the relationship between LSBMD Z-scores and VF varies depending on the LSBMD reference databases used to generate the Z-scores.

Subjects and Methods

Subjects and clinical data

Patients were recruited through pediatric oncology clinics in 10 children’s hospitals across Canada as part of the STeroid-associated Osteoporosis in the Pediatric Population (STOPP) research program (6, 7, 13). The study was approved by the Ethics Board in each institution, and informed consent/assent was obtained, as appropriate. Children from 1 month to 17 years of age with ALL were enrolled from 2005 to 2007 with the baseline bone health assessment targeted within 30 days of glucocorticoid therapy initiation (median 20 days; interquartile range (IQR) 11 to 26 days). Among the 186 children enrolled in the study, 108 (58%) were boys and 140 (75%) were Caucasian. The ethnicities for the other 25% of the cohort were as follows: native/aboriginal people of North America (North American Indian, Métis, Inuit/Eskimo) (7%), South Asian (eg, East Indian, Pakistani, Punjabi, Sri Lankan) (5%), black (4%), and other (9%).

Standard demographic data including age, gender, and ethnicity were recorded (6). Children were treated according to the Children’s Oncology Group (6) (9 sites) or the Dana-Farber Cancer Institute (1 site) protocols. A comprehensive clinical description of this cohort has been published in 2 reports elsewhere (6, 7).

BMD (grams per square centimeter) was measured in the anterior-posterior projection at the LS (L1–L4) by DXA using cross-calibrated Hologic or Lunar Prodigy (GE Lunar Corporation) systems to generate an areal LSBMD raw value. Details on Hologic and Lunar machine cross-calibration methodology have been described in previous reports (6, 7). All scans were analyzed centrally by 1 technologist using Hologic QDR (version 12.0) software.

Lateral thoracolumbar spine radiographs were scored independently by 2 pediatric radiologists according to the modified Genant semiquantitative method (14), and a third radiologist adjudicated discrepancies. Vertebral bodies were graded according to the extent of the difference in height ratios from 100% when the anterior vertebral height was compared with the posterior height, the middle height to the posterior height, and the
posterior height to the posterior height of adjacent vertebral bod-
ies, with the following grade definitions: grade 0 (normal), 20% or less; grade 1 (mild), more than 20% to 25%; grade 2 (mod-
erate), more than 25% to 40%; grade 3 (severe), more than 40%.
Grade 1 or higher scores were considered to represent prevalent
VF.

Pediatric reference data for generation of LSBMD
Z-scores

A literature search was performed for pediatric LSBMD ref-
ere data that had been generated using either Hologic or Lu-
nar machines. These 2 machines were chosen because they are
currently used across Canada and are the most commonly em-
ployed machines as reported in the pediatric literature. Norma-
tive reference databases that satisfied the following criteria were
included in this study: 1) comprised of primarily Caucasian chil-
dren because this was the composition of our cohort; 2) reference
data that spanned a minimum age range of 5 years to less than 18
years or reference data for infants and toddlers (0 to 3 years); 3)
sufficient methodological detail to permit calculation of age- and
gender-specific LSBMD Z-scores; and 4) published in English. In
addition, the Hologic Apex 3.1 (15) database, an unpublished
database generated by the Hologic manufacturer, was also in-
cluded in this study given its frequent use in clinical practice. In
some cases, related databases were combined to generate a single
database that spanned the age relevant to this pediatric ALL
cohort (0–18 years).

LSBMD Z-score calculations

LSBMD Z-scores (for either L1–L4 or L2–L4, depending
upon the parameter that was given by the normative reference
databases) were calculated using the selected published normative
reference databases. The databases provided either LMS param-
eters (the power in the Box–Cox transformation (L), the median
(M), the generalized coefficient of variation (S)), mean and stan-
dard deviation (M/SD), parameters, or gender- and age-specific
formulae. Linear interpolation was applied to determine Z-scores
for ages that were intermediate to age-related values provided in
the normative reference databases. Ethnicity was taken into ac-
count for the LSBMD Z-score calculation when race-specific
reference databases were available. Specifically, black children
were analyzed on black reference data where available, and Cau-
savian/nonblack reference data were applied to Caucasian chil-
dren. For the other ethnicities in this cohort, Caucasian/nonblack
reference data were applied to generate the LSBMD Z-scores due
to lack of race-specific data.

Based on the age ranges available in the various normative
reference databases, we defined six age groups for comparison of
LSBMD Z-scores: 1) 0 to 18 years (n = 186; 108 (58%) male; mean
± SD age 6.6 ± 4.0 years; 29 (16%) with VF); 2) 0 to 5
years (n = 86; 53 (62%) male; age 3.3 ± 0.9 years; 10 (12%) with
fractures); 3) 3 to 18 years (n = 153; 87 (57%) male; age 7.5 ±
3.9 years; 26 (17%) with fractures); 4) 5 to 18 years (n = 100;
56 (56%) male; age 9.4 ± 3.5 years; 19 (19%) with fractures);
5) 8 to 16 years (n = 51; 26 (51%) male; age 11.8 ± 2.1 year; 10
(20%) with fractures); and 6) 9 to 14 years (n = 37; 19 (51%)
males; age 11.7 ± 1.5 years; 7 (19%) with fractures).

Height Z-score-adjusted LSBMD Z-scores for children in the
groups aged 5 to 18 years, 8 to 16 years, and 9 to 14 years were
also calculated, according to the method and reference data pro-
vided by Zemel et al (16). The height adjustment was limited to
these 3 different age ranges, because this was the age range avail-
able in the Zemel publication. Note that the original reference
data published by Zemel et al (16) was subsequently revised
using updated software (17); this revised data was used to cal-
culate the height Z-score-adjusted LSBMD Z-scores in the cur-
rent report (17).

Statistical analysis

LSBMD Z-scores generated by various reference databases
were reported as means and SD and were compared using the
Friedman test. Pearson correlation coefficients (r) between
LSBMD Z-scores generated using any pair of normative refer-
dence databases were calculated. Based on the cutoff threshold
of −2, LSBMD Z-scores were categorized as low (≤−2) or normal
(>−2) (12). A logistic regression model with VF as the outcome
and LSBMD Z-score as the continuous risk factor was developed
each normative reference database. The model was adjusted
for height and weight Z-scores when LSBMD Z-score was used
in the model. The model was adjusted for weight Z-score only,
when the height-adjusted LSBMD Z-score was used.

The odds ratios (OR) expressed the risk of VF for every 1 SD
decrease in LSBMD Z-score. The area under the receiver oper-
ating characteristic curve (AUC), which measures how well
LSBMD Z-scores can distinguish between children with and
without VF, was determined. Cochran’s Q test (18) was con-
ducted to assess the heterogeneity/variation in the OR and AUC
within each age group across different normative reference da-
tabases. Continuous net reclassification improvement (NRI)
(19) was calculated to assess the improvement in prediction of VF
risk for every 1 SD decrease in LSBMD Z-score gained by using
any other normative reference database in comparison with Ho-
logic version 12.3 (chosen as the reference because this is the
normative database used in the STOPP study) (20). Analyses
were conducted using SAS version 9.3 (SAS Institute Inc). Sta-
tistical significance was considered at P < .05.

Results

Vertebral fractures

There were a total of 75 VF (53 thoracic, 22 lumbar) in
29 of 186 patients (16%), as previously described (6). Of
the children with VF, 48% had mild fractures as the worst
grade, and 52% had moderate or severe VF. Additional
details on the distribution of VF that were found in this
cohort of children with ALL around the time of diagnosis
have been previously published (6). There was no signif-
cant difference in the mean (SD) spine areal BMD raw
scores (after cross-calibration) for those children with
lumbar fractures (n = 12) when the fractured vertebrae
were included in the L1 to L4 BMD results (0.43 g/cm^2
(0.14)) compared with when the fractured vertebrae were
excluded (0.44 g/cm^2 (0.13), P = .75 according to an in-
dependent 2-sample t test).

Correlations among LSBMD Z-scores derived on
different normative reference databases

Our literature search yielded 15 databases that met
the criteria outlined in the Subjects and Methods. Table
Table 1. Published Pediatric LSBMD Normative Reference Databases

<table>
<thead>
<tr>
<th>Reference</th>
<th>Manufacturer and Machine</th>
<th>Software</th>
<th>Country, Year</th>
<th>Age Range, y</th>
<th>Subjects (Female/Male)</th>
<th>Study Design</th>
<th>Skeletal Site</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hologic 12.3 (20)</td>
<td>Hologic</td>
<td>12.3</td>
<td>NA</td>
<td>1–18</td>
<td>NA</td>
<td>NA</td>
<td>L1–4</td>
<td>M/SD</td>
</tr>
<tr>
<td>Hologic Apex 3.1 (15)</td>
<td>Hologic</td>
<td>3.1</td>
<td>NA</td>
<td>3–85</td>
<td>NA</td>
<td>NA</td>
<td>L1–4</td>
<td>LMS</td>
</tr>
<tr>
<td>Kalkwarf 2013 (21)</td>
<td>Hologic Discovery A</td>
<td>12.7</td>
<td>USA</td>
<td>Argentina, Australia, Northern Europe</td>
<td>2010–2011</td>
<td>1207 (752/455)</td>
<td>SP</td>
<td>L1–4</td>
</tr>
<tr>
<td>Kalkwarf 2013 (21)</td>
<td>Hologic Discovery A</td>
<td>8.26, 12.3</td>
<td>USA</td>
<td>2010–2011</td>
<td>1207 (752/455)</td>
<td>SP</td>
<td>L1–4</td>
<td>Formula</td>
</tr>
</tbody>
</table>

* P represents prospective design, X cross-sectional design, S single-center study, and M multicenter study.

1. Shows the age ranges encompassed by each of these reference data, the nationalities of the children, and the parameters/calculations used for Z-score generation. Two additional normative reference databases encompassing the full pediatric age range (0–18 years) were constructed by combining reference data from different sources. The first was compiled by combining the Hologic Apex 3.1 database (15) with the published Hologic data from Kalkwarf et al (21). The second was constructed by combining the published Hologic data from Zemel et al (16) with those from Kalkwarf et al (21). These databases were combined to allow evaluation of the entire study population, given the similarity in methodological approaches between the 2 databases and the paucity of published databases covering the age range of our cohort.

Correlations of LSBMD Z-scores generated on each pair of normative reference databases within each age group ranged from \( r = 0.90 \) to 0.99 for the age group 0 to 18 years, 0.95 to 0.99 for the age group 3 to 18 years, 0.90 to 0.99 for the age group 5 to 18 years, 0.88 to 0.99 for the age group 8 to 16 years, 0.85 to 0.99 for the age group 9 to 14 years, and \( r = 0.92 \) for the 0 to 5 years group. Most coefficients (94%) were above 0.90. The lowest were 0.85 between the Hologic 12.3 (20) and Bachrach 1999 (22) in the age group 9 to 14 years. The highest correlation coefficients indicated a strong linear relationship between the LSBMD Z-scores generated using different normative reference databases.

The mean LSBMD Z-scores generated by the normative reference databases varied from −2.0 to −0.5 for the age group 0 to 18 years, −2.0 to −1.4 for the age group 0 to 5 years −2.1 to −0.3 for the age group 3 to 18 years, −2.0 to −0.1 for the age group 5 to 18 years, −1.9 to 0.1 for the age group 8 to 16 years, and −2.0 to 0 for the age group 9 to 14 years (Table 2). The mean LSBMD Z-scores were ≤0 for all databases, with the exception of the Webber database (23) (used for age group 8–16 years and age group 9–14 years). The Del Rio (24) normative reference database produced the lowest mean Z-scores within each age group. Overall, the mean LSBMD Z-scores were significantly different within each age group for the various reference databases (\( P < .001 \)). The maximum disparity in the mean LSBMD Z-score was 2.0 SD (between the Webber (23) and the Del Rio (24) databases for the age groups 8–16 years and 9–14 years).

**Proportion of children assigned LSBMD Z-scores ≤−2 SD**

Overall, the numbers and percentages of children with LSBMD Z-scores at or below −2.0 SD differed substantially depending upon the database used for the Z-score calculation (Table 2). Among the entire cohort of 186 children with ALL from 0 to 18 years of age, the discrepancy between databases ranged from 15% of children ≤−2.0 SD for the Hologic Apex 3.1 (15) and Kalkwarf (21) combined database, to 48% of children for the Del Rio database (24). Among the normative reference databases studying children 5 to 18 years of age, the Del Rio database (24) designated spine BMD Z-score at or below −2.0 in 53% of the study subjects compared with 7% on the Webber database (23).

Among the 29 children with ALL from 0 to 18 years of age who had VF, the percentages of children with LSBMD
Z-scores ≤ −2.0 SD on the various databases were as follows: 52% for the Hologic 12.3 database (20), 38% for the Hologic Apex 3.1 (15) and Kalkwarf (21) combined database, 38% for the Zemel (16) and Kalkwarf (21) combined database, and 66% for the Del Rio database (24). Among the 19 children in the 5 to 18 years age group who had VF, the proportion with height-adjusted LSBMD Z-scores worse than −2 ranged from 21% for the Webber (23) and Zemel databases (16) to 74% for the van der Sluis database (25) (Figure 1).

### Table 2. Proportion of Study Subjects Assigned LSBMD Z-scores ≤ −2

<table>
<thead>
<tr>
<th>Age Group (Proportion of Children With Fractures)</th>
<th>Reference Database</th>
<th>LSBMD Z-Score Mean (SD)</th>
<th>LSBMD Z-Scores ≤ −2 n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–18 y (29/186)</td>
<td>Hologic 12.3 (20)</td>
<td>−1.2 (1.3)</td>
<td>46 (25)</td>
</tr>
<tr>
<td></td>
<td>Hologic Apex 3.1 (15) + Kalkwarf 2013 (21)</td>
<td>−0.5 (1.5)</td>
<td>28 (15)</td>
</tr>
<tr>
<td></td>
<td>Zemel 2011 (16) + Kalkwarf 2013 (21)</td>
<td>−0.6 (1.6)</td>
<td>31 (17)</td>
</tr>
<tr>
<td></td>
<td>Del Rio 1994 (24)</td>
<td>−0.2 (1.1)</td>
<td>30 (16)</td>
</tr>
<tr>
<td>0–5 y (10/86)</td>
<td>Hologic 12.3 (20)</td>
<td>−1.4 (1.2)</td>
<td>22 (26)</td>
</tr>
<tr>
<td></td>
<td>Del Rio 1994 (24)</td>
<td>−2.0 (1.0)</td>
<td>37 (43)</td>
</tr>
<tr>
<td>3–18 y (26/153)</td>
<td>Hologic 12.3 (20)</td>
<td>−1.2 (1.3)</td>
<td>39 (25)</td>
</tr>
<tr>
<td></td>
<td>Hologic Apex 3.1 (15)</td>
<td>−0.5 (1.5)</td>
<td>21 (14)</td>
</tr>
<tr>
<td></td>
<td>Del Rio 1994 (24)</td>
<td>−2.1 (1.2)</td>
<td>81 (53)</td>
</tr>
<tr>
<td></td>
<td>Kelly 2005 (34)</td>
<td>−0.6 (1.4)</td>
<td>21 (14)</td>
</tr>
<tr>
<td></td>
<td>Webber 2007 (23)</td>
<td>−0.3 (1.3)</td>
<td>15 (10)</td>
</tr>
<tr>
<td>5–18 y (19/100)</td>
<td>Hologic 12.3 (20)</td>
<td>−1.0 (1.3)</td>
<td>24 (24)</td>
</tr>
<tr>
<td></td>
<td>Hologic Apex 3.1 (15)</td>
<td>−0.3 (1.4)</td>
<td>9 (9)</td>
</tr>
<tr>
<td></td>
<td>Zemel 2011 (16)</td>
<td>−0.3 (1.4)</td>
<td>9 (9)</td>
</tr>
<tr>
<td></td>
<td>Del Rio 1994 (24)</td>
<td>−2.0 (1.2)</td>
<td>53 (53)</td>
</tr>
<tr>
<td></td>
<td>Kelly 2005 (34)</td>
<td>−0.4 (1.3)</td>
<td>9 (9)</td>
</tr>
<tr>
<td></td>
<td>Del Rio 1994 (24)</td>
<td>−1.4 (1.1)</td>
<td>29 (29)</td>
</tr>
<tr>
<td></td>
<td>van der Sluis 2002 (25)</td>
<td>−1.9 (2.2)</td>
<td>50 (50)</td>
</tr>
<tr>
<td></td>
<td>Webber 2007 (23)</td>
<td>−0.1 (1.3)</td>
<td>7 (7)</td>
</tr>
<tr>
<td>8–16 y (10/51)</td>
<td>Hologic 12.3 (20)</td>
<td>−0.8 (1.4)</td>
<td>9 (18)</td>
</tr>
<tr>
<td></td>
<td>Hologic Apex 3.1 (15)</td>
<td>−0.1 (1.3)</td>
<td>3 (6)</td>
</tr>
<tr>
<td></td>
<td>Zemel 2011 (16)</td>
<td>−0.1 (1.2)</td>
<td>3 (6)</td>
</tr>
<tr>
<td></td>
<td>Del Rio 1994 (24)</td>
<td>−0.4 (1.3)</td>
<td>5 (10)</td>
</tr>
<tr>
<td></td>
<td>Falkner 1996 (28)</td>
<td>−0.2 (1.5)</td>
<td>4 (8)</td>
</tr>
<tr>
<td></td>
<td>Kalkwarf 2007 (33)</td>
<td>−0.1 (1.3)</td>
<td>4 (8)</td>
</tr>
<tr>
<td></td>
<td>Kelly 2005 (34)</td>
<td>−0.3 (1.2)</td>
<td>3 (6)</td>
</tr>
<tr>
<td></td>
<td>Del Rio 1994 (24)</td>
<td>−1.4 (1.2)</td>
<td>14 (27)</td>
</tr>
<tr>
<td></td>
<td>van der Sluis 2002 (25)</td>
<td>−1.6 (1.1)</td>
<td>17 (33)</td>
</tr>
<tr>
<td></td>
<td>Webber 2007 (23)</td>
<td>0.1 (1.3)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>9–14 y (7/37)</td>
<td>Hologic 12.3 (20)</td>
<td>−0.9 (1.3)</td>
<td>8 (22)</td>
</tr>
<tr>
<td></td>
<td>Hologic Apex 3.1 (15)</td>
<td>−0.2 (1.2)</td>
<td>2 (5)</td>
</tr>
<tr>
<td></td>
<td>Zemel 2011 (16)</td>
<td>−0.2 (1.2)</td>
<td>2 (5)</td>
</tr>
<tr>
<td></td>
<td>Del Rio 1994 (24)</td>
<td>−0.5 (1.2)</td>
<td>4 (11)</td>
</tr>
<tr>
<td></td>
<td>Bachrach 1999 (22)</td>
<td>−1.2 (1.1)</td>
<td>9 (24)</td>
</tr>
<tr>
<td></td>
<td>Faulkner 1996 (28)</td>
<td>−0.3 (1.5)</td>
<td>3 (8)</td>
</tr>
<tr>
<td></td>
<td>Fonseca 2001 (32)</td>
<td>−1.7 (1.3)</td>
<td>14 (38)</td>
</tr>
<tr>
<td></td>
<td>Kalkwarf 2007 (33)</td>
<td>−0.2 (1.2)</td>
<td>3 (8)</td>
</tr>
<tr>
<td></td>
<td>Kelly 2005 (34)</td>
<td>−0.4 (1.1)</td>
<td>2 (5)</td>
</tr>
<tr>
<td></td>
<td>Pludowski 2005 (26)</td>
<td>−1.5 (1.1)</td>
<td>11 (30)</td>
</tr>
<tr>
<td></td>
<td>van der Sluis 2002 (25)</td>
<td>−1.7 (1.0)</td>
<td>13 (35)</td>
</tr>
<tr>
<td></td>
<td>Webber 2007 (23)</td>
<td>0.0 (1.3)</td>
<td>1 (3)</td>
</tr>
</tbody>
</table>

*Summary statistics were calculated using height Z-score-adjusted LSBMD Z-score for this age group based on the ages available for this calculation according to Zemel et al (16).*

### Association between LSBMD Z-score and vertebral fracture

For the age group 0 to 18 years, OR for the association between LSBMD Z-score and VF were similar and ranged from 1.92 (95% confidence interval (CI), 1.44, 2.56) to 2.70 (95% CI, 1.70, 4.28) per SD decrease in Z-score (Table 3); the nonsignificant *P* value from the Cochran’s Q test confirmed that these differences were statistically insignificant. These results also confirmed that lower LSBMD Z-scores were significantly associated with in-
creased risk of VF. Results for other age groups showed similar patterns, ie, similar OR for the relationship between LSBMD Z-scores and VF, regardless of the normative database used to generate the Z-scores (Table 3).

The predictive accuracy of LSBMD Z-scores for VF was tightly clustered across different normative reference databases for each age group (Table 3). For example, the AUC for the age group 0 to 18 years ranged from 0.71 (95% CI, 0.60–0.83) to 0.75 (95% CI, 0.63–0.86). Likewise, similar AUC results were present in each age group, confirmed by the nonsignificant P values from the Cochran’s Q test. Similarly, there was no significant improvement in NRI by any normative reference database compared with Hologic 12.3 (20) except for the Del Rio (24) reference database for the age group 0 to 5 years, Hologic Apex 3.1 (15) and Pludowski (26) reference databases for the age group 8 to 16 years, and Zemel (16), Lunar (USA) (27), Bachrach (22), Faulkner (28), and van der Sluis (25) reference databases for the age group 9 to 14 years. These exceptions may be due to 1) the inflated differences in predictive accuracies associated with the NRI; and/or 2) unstable estimates arising from the logistic regression models, given that the overall sample size and the number of children with VF in these 3 age groups were small. For the age group 0 to 18 years (which spans the age range in the entire leukemia STOPP cohort), the NRI were similar compared with Hologic 12.3 (20) for all of the studied reference databases, ranging from −0.15 (95% CI, −0.54, 0.25; P = .46) to 0.07 (95% CI, −0.32, 0.47; P = .72).

Discussion

We found that in our cohort of children with an acute critical illness and the potential for significant bone morbidity, the mean LSBMD Z-scores varied substantially with maximum disparity of 1.5, 0.6, 1.8, 1.9, 2.0, and 2.0 SD for age groups 0 to 18, 0 to 5, 3 to 18, 5 to 18, 8 to 16, and 9 to 14 years, respectively (Table 2). Not surprisingly, the LSBMD Z-scores arising from the largest normative reference databases (ie, those on more than 600 children, which include normative reference databases generated by the machine manufacturers) gave midrange results overall and were more similar to each other, whereas the greatest disparities arose from normative databases generated on small numbers of children. This observed variability likely results from a host of factors, including differences in the brand, model, and software used for LSBMD acquisition, as well as inherent differences in the sample populations upon which the normative databases were derived (arising from genetic and/or lifestyle differences).

The consequences of a single absolute BMD value generating markedly different Z-scores depending on the normative reference database used are important in 2 areas: 1) clinical practice, where the Z-score has traditionally been part of assessing a child's skeletal status and defining the presence or absence of osteoporosis, and 2) research studies, which attempt to define the relationship between BMD and other clinical risk factors and fracture outcomes.

For the clinician evaluating bone health in a child with or at risk for a VF (ie, at the individual patient level), the use of LSBMD Z-scores is problematic given the disparate Z-scores that are generated by the various normative reference databases. In our study population, the proportion falling below the traditionally diagnostic BMD threshold of ≤−2 (who would be categorized as below the expected range for age), varied markedly for children 0 to 18 years of age, from 15% for the Hologic Apex 3.1 (15) and Kalkwarf (21) combined database, to 48% using the Del Rio (24) database.

The inaugural (2008) ISCD criteria for the diagnosis of osteoporosis in children included a BMD Z-score ≤−2 as part of the definition (29). However, updated guidelines published in 2014 state that the presence of 1 or more low-trauma VF is sufficient to diagnose osteoporosis, regardless of BMD findings (12). Our results provide concrete evidence to support this new definition, because we have shown that children with ALL and VF often have spine BMD Z-scores above −2.0 SD (although with variability, depending upon the normative database used to generate the Z-scores). Specifically, among the 29 children with leukemia in our cohort and with VF, the proportion...
there was a 15% increased non-VF risk. In line with that for every 1 SD reduction in distal femur BMD by Henderson et al (9) showed in children with cerebral palsy, the fracture history in the non-VF setting. Interestingly, fracture risk. This proviso emphasizes the importance of caveat in the new ISCD guidelines to the effect that a BMD trauma, non-VF) (12). As part of this scenario, there is a of the definition of osteoporosis in children who have clin- depending on the normative database that was used.

assigned a LSBMD Z-score better than –2 SD (ie, in the face of overt bone fragility) ranged from 38% to 66% depending on the normative database that was used.

At the same time, the new (2014) ISCD criteria have retained a BMD Z-score threshold at or below –2 as part of the definition of osteoporosis in children who have clinically significant fractures other than at the spine (ie, low-trauma, non-VF) (12). As part of this scenario, there is a caveat in the new ISCD guidelines to the effect that a BMD or bone mineral content Z-score ≤ –2.0 in a child with a non-VF does not preclude skeletal fragility and increased fracture risk. This proviso emphasizes the importance of the fracture history in the non-VF setting. Interestingly, Henderson et al (9) showed in children with cerebral palsy, that for every 1 SD reduction in distal femur BMD by DXA, there was a 15% increased non-VF risk. In line with our results, Henderson et al (9) also reported that whereas the prevalence of non-VF increased with lower distal femur Z-scores, some children nevertheless had non-VF despite distal femur BMD Z-scores > –1 SD. These data in children with cerebral palsy also support an osteoporosis definition that emphasizes the fracture history in those with non-VF.

Ultimately, what may bypass these issues is a standard-ized LSBMD normative reference database generated using a representative sample of children who are scanned on both Hologic and Lunar machines. When such a database is available, studies could then be done to test clinically relevant Z-score thresholds, which might differ from the arbitrary value of –2.0 that has been used to this point. The incorporation of other risk factors beyond BMD into fracture prediction models (such as has been done for

<table>
<thead>
<tr>
<th>Age Group (proportion of children with fractures)</th>
<th>Reference Database</th>
<th>OR* (95% CI)</th>
<th>P Value From Cochran’s Q Test</th>
<th>AUC (95% CI)</th>
<th>OR* (95% CI)</th>
<th>P Value From Cochran’s Q Test</th>
<th>NRI (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–18 y (29/166)</td>
<td>Hologic 12.3 (20)</td>
<td>2.36 (1.58, 3.53)</td>
<td>.61 (0.74, 0.83)</td>
<td>.98</td>
<td>Reference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–5 y (10/86)</td>
<td>Hologic 12.3 (20)</td>
<td>11.36 (2.78, 46.40)</td>
<td>.98 (0.95, 1.00)</td>
<td>.61</td>
<td>Reference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3–18 y (26/153)</td>
<td>Hologic 12.3 (20)</td>
<td>2.21 (1.47, 3.35)</td>
<td>.99 (0.73, 0.80)</td>
<td>.80</td>
<td>&gt;.99 Reference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5–18 y (19/100)</td>
<td>Hologic 12.3 (20)</td>
<td>1.65 (1.07, 2.55)</td>
<td>&gt;.99 (0.67, 0.82)</td>
<td>&gt;.99</td>
<td>Reference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8–16 y (10/51)</td>
<td>Hologic 12.3 (20)</td>
<td>2.31 (1.16, 4.59)</td>
<td>&gt;.99 (0.76, 0.70)</td>
<td>&gt;.99</td>
<td>Reference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9–14 y (7/37)</td>
<td>Hologic 12.3 (20)</td>
<td>4.27 (1.36, 13.44)</td>
<td>.97 (0.85, 1.60)</td>
<td>.97</td>
<td>&gt;.99 Reference</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* OR are interpreted as an increased risk of vertebral fracture per SD decrease in LSBMD Z-score.

** a Statistical analyses are based on the height Z-score-adjusted LSBMD Z-scores for this age group.

<table>
<thead>
<tr>
<th>Age Group (proportion of children with fractures)</th>
<th>Reference Database</th>
<th>OR* (95% CI)</th>
<th>P Value From Cochran’s Q Test</th>
<th>AUC (95% CI)</th>
<th>OR* (95% CI)</th>
<th>P Value From Cochran’s Q Test</th>
<th>NRI (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–18 y (29/166)</td>
<td>Hologic 12.3 (20)</td>
<td>2.36 (1.58, 3.53)</td>
<td>.61 (0.74, 0.83)</td>
<td>.98</td>
<td>Reference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–5 y (10/86)</td>
<td>Hologic 12.3 (20)</td>
<td>11.36 (2.78, 46.40)</td>
<td>.98 (0.95, 1.00)</td>
<td>.61</td>
<td>Reference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3–18 y (26/153)</td>
<td>Hologic 12.3 (20)</td>
<td>2.21 (1.47, 3.35)</td>
<td>.99 (0.73, 0.80)</td>
<td>.80</td>
<td>&gt;.99 Reference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5–18 y (19/100)</td>
<td>Hologic 12.3 (20)</td>
<td>1.65 (1.07, 2.55)</td>
<td>&gt;.99 (0.67, 0.82)</td>
<td>&gt;.99</td>
<td>Reference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8–16 y (10/51)</td>
<td>Hologic 12.3 (20)</td>
<td>2.31 (1.16, 4.59)</td>
<td>&gt;.99 (0.76, 0.70)</td>
<td>&gt;.99</td>
<td>Reference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9–14 y (7/37)</td>
<td>Hologic 12.3 (20)</td>
<td>4.27 (1.36, 13.44)</td>
<td>.97 (0.85, 1.60)</td>
<td>.97</td>
<td>&gt;.99 Reference</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* OR are interpreted as an increased risk of vertebral fracture per SD decrease in LSBMD Z-score.

** a Statistical analyses are based on the height Z-score-adjusted LSBMD Z-scores for this age group.
adults, FRAX) (30) will also likely strengthen the ability to predict VF and non-VF in children.

There are a number of strengths and limitations to this study. This report is novel in its exploration of the disparity in BMD Z-scores generated by different normative databases, first by including both Hologic and Lunar machines and second by studying the relationship between LSBMD Z-scores and VF depending on the normative database that is used to generate the Z-scores. In addition, this report provides data on a cohort that is representative of children who are frequently referred for a bone health evaluation, and so places the issues facing the clinician into sharp focus. Furthermore, calculation of the AUC and NRI provides the clinician with the opportunity to assess the predictive accuracy of LSBMD Z-scores for VF children depending on the normative database that is chosen for implementation in the clinic. A limitation of this study is that only the relationship between LSBMD and VF was explored, and not the relationship between BMD and other fracture types. Another is that only spine BMD was assessed, and not BMD at other skeletal sites. A third limitation is that our study subjects all have ALL; additional studies will be needed to determine whether similar relationships exist in other disease types. As well, we acknowledge that the differences in machine technology (ie, pencil vs fan beam scan acquisition for the various reference databases) may have influenced the accuracy of the data. At the same time, Faulkner et al (31) showed that fan vs pencil beam technology did not lead to clinically significant differences.

In summary, we observed marked disparity in LSBMD Z-scores depending on the normative database that was used, and that children with VF frequently had LSBMD Z-scores >-2 SD. On the other hand, the relationship between LSBMD Z-scores and VF was consistent among the different normative reference databases. These results suggest that although the use of a LSBMD Z-score threshold as part of the definition of osteoporosis in a child with VF does not appear valid, the study of relationships between BMD and VF remains valid regardless of the BMD database that is used.

Acknowledgments

The Canadian STOPP Consortium, a national pediatric bone health working group, would like to thank the children and their families who participated in the study and without whom the STOPP study would not have been possible. We also thank the research associates who managed the study at the coordinating center (the Children’s Hospital of Eastern Ontario Ottawa, Ontario); Elizabeth Sykes (STOPP Project Manager), Maya Scharke (STOPP Data Analyst and Database Manager), Monica Tomiak (Statistical Analyses), Victor Konji (STOPP Publications and Presentations Committee Liaison), Steve Anderson (Children’s Hospital of Eastern Ontario Pediatric Bone Health Program Research Manager), Catherine Riddell (STOPP National Study Monitor); research associates who took care of the patients from the following institutions: Alberta Children’s Hospital, Calgary, Alberta: Eileen Pyra; British Columbia Children’s Hospital, Vancouver British Columbia: Terry Viczko, Sandy Hwang, Angelyne Sarmiento; Children’s Hospital of Eastern Ontario, Ottawa, Ontario: Heather Cosgrove, Josie MacLennan, Catherine Riddell; Children’s Hospital, London Health Sciences Centre, London, Ontario: Vinolia Arthur Hayward, Leila MacBean, Mala Ramu; McMaster Children’s Hospital, Hamilton, Ontario: Susan Docherty-Skippen; IWK Health Center, Halifax, Nova Scotia: Cindy Campbell, Aleasha Warner; Montréal Children’s Hospital, Montréal, Québec: Valérie Gagné, Diane Laforte, Maritza Laprise; Ste. Justine Hospital, Montréal, Québec: Claude Belleville, Natacha GaulinMarion; Stollery Children’s Hospital, Edmonton, Alberta: Ronda Blasco, Germaine MelUne, Amanda Mullins; Toronto Hospital for Sick Children, Toronto, Ontario: Alexandra Airhart, Michele Petrovic, Nicole Sarvaria; Winnipeg Children’s Hospital, Winnipeg, Manitoba: Dan Catte, Erika Bloomfield, Jeannine Schellenberg; and the research nurses, support staff, and all the STOPP collaborators from the various Divisions of Nephrology, Oncology, Rheumatology, and Radiology who have contributed to the care of the children enrolled in the study.

Address all correspondence and requests for reprints to: Dr Jinhui Ma, PhD, Associate Scientist, Children’s Hospital of Eastern Ontario Research Institute, 401 Smyth Road, Ottawa, Ontario, Canada. E-mail: jma@cheo.on.ca.

This study was primarily funded by an operating grant from the Canadian Institutes for Health Research (FRN 64285). Additional funding for this work has been provided to L.W. by the Canadian Institutes for Health Research New Investigator Program, the Canadian Child Health Clinician Scientist Career Enhancement Program, a University of Ottawa Research Chair Award and the CHEO Departments of Pediatrics and Surgery. This work was also supported by the Children’s Hospital of Eastern Ontario Research Institute and the University of Alberta Women and Children’s Health Research Institute.

The Canadian Steroid-associated Osteoporosis in the Pediatric Population (STOPP) Consortium (a pan-Canadian, pediatric bone health working group): *Principal Investigator; +Executive Committee Member; $Publications and Presentations Committee Member.

Coordinating Center: Children’s Hospital of Eastern Ontario, Ottawa, ON: Leanne M. Ward*+ ($Study Principal Investigator), Janusz Feber*+ ($Nephrology), Jacqueline Halton*+ ($Oncology), Roman Jurencak ($Rheumatology), MaryAnn Matzinger ($Radiology, Central Radiograph Analyses), Johannes Roth ($Rheumatology), Nazih Shenouda* ($Radiology, Central Radiograph Analyses), Jinhui Ma ($Research Methods and Statistics), Ottawa Hospital Research Institute, Ottawa Methods Centre Ottawa, ON: David Moher*+ ($Research Methods), Monica Taldjaar ($Research Methods and Statistics).

Participating Centers: Alberta Children’s Hospital, Calgary, AB: Josephine Ho (Site Principal Investigator, from Jul 2013 to
current), David Stephure (Bone Health, Site Principal Investigator until Jul, 2013), Reinhard Kloiber (Radiology), Victor Lewis (Oncology), Julian Midgley (Nephrology), Paivi Miettunen (Rheumatology), BC Children’s Hospital, Vancouver, BC; David Cabral* (Site Principal Investigator), David B. Dix (Oncology), Kristin Houghton (Rheumatology), Helen R. Nadel (Radiology), BC Women’s Health Sciences Centre, and Department of Radiology, University of British Columbia, Vancouver, BC: Brian C. Lentle§ (Radiology), Brock University, Faculty of Applied Health Sciences, St. Catharines, ON: John Hay§ (Physical Activity Measurements), Children’s Hospital, London Health Sciences Centre, University of Western Ontario, London, ON: Robert Stein (Site Principal Investigator), Elizabeth Cairney (Oncology), Cheryl Clarison (Bone Health), Guido Filler (Nephrology)§, Joanne Grimmer (Nephrology), Scott McKillop (Radiology), Tom Blydt-Hansen (Nephrology), Sara Israels (Oncology), Kiem Oen (Rheumatology), Martin Reed (Radiology), Celia Rodd§ (Site Principal Investigator), Josée Dubois (Radiology), Caroline Laverdère (Oncology), Véronique Phan (Nephrology), Claire Saint-Cyr (Rheumatology), Stollery Children’s Hospital, Edmonton, AB: Robert Couch*§ (Site Principal Investigator), Janet Ellsworth (Rheumatology), Maury Pink (Nephrology), Jacob Jaremko and Kerry Siminoski§ (Radiology), Beverly Wilson (Oncology), Toronto Hospital for Sick Children, Montréal, Québec: Anne Marie Sbrocchi (Site Principal Investigator, from 2013 to current), Celia Rodd§ (Site Principal Investigator, until 2013), Sharon Abish (Oncology), Lorraine Bell (Nephrology), Claire LeBlanc (Rheumatology), Rosie Scuccimarrì (Radiology), Shriners Hospital for Children, Montréal, Québec: Frank Rauch§* (Co-Chair, Publications and Presentations Committee and Ancillary Studies Committee), Ste. Justine Hospital, Montréal, Québec: Nathalie Alos*§ (Site Principal Investigator), Josée Dubois (Radiology), Caroline Laverdère (Oncology), Véronique Phan (Nephrology), Claire Saint-Cyr (Rheumatology), Stollery Children’s Hospital, Edmonton, AB: Robert Couch*§ (Site Principal Investigator), Janet Ellsworth (Rheumatology), Maury Pink (Nephrology), Jacob Jaremko and Kerry Siminoski§ (Radiology), Beverly Wilson (Oncology), Toronto Hospital for Sick Children, Montréal, ON: Ronald Grant* (Site Principal Investigator), Martin Charron (Radiology, until 2013), Diane Hebert (Nephrology), Université de Sherbrooke, Department of family medicine, Sherbrooke, Québec: Isabelle Gaboury§ (Biostatistics), Winnipeg Children’s Hospital, Winnipeg, MB: Shayne Taback§ (Site Principal Investigator), Tom Blydt-Hansen (Nephrology), Sara Israels (Oncology), Kiem Oen (Rheumatology), Martin Reed (Radiology), Celia Rodd§ (Bone Health, from 2013 to current).

Disclosure Summary: The authors have nothing to disclose.

References


