Original Research Article

Evaluating the predictive quality of the Chapman bone algorithm using aggregated data sets

Noah D. Barrett¹, Cameron W. James¹, Joshua P. Tam¹, Elise S. Levesque², Anton S. Ketterer², Wajiha R. Memon¹, Cyril S. Rakovski³, Frank Frisch¹*

¹Department of Health Sciences, ²Department of Biological Sciences, ³Department of Mathematics, Chapman University, Orange, CA, USA

Received: 17 October 2018
Accepted: 30 November 2018

*Correspondence:
Dr. Frank Frisch,
E-mail: frisch@chapman.edu

Copyright: © the author(s), publisher and licensee Medip Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Background: Due to an aging population, osteoporosis has become an increasingly prevalent metabolic bone disorder that is largelyundiagnosed worldwide because of inaccessible and expensive DXA machines. The Chapman bone algorithm (CBA), a mathematical treatment that enables osteoporosis determination by using simply-assayed bone metabolites from blood serum, has been previously presented as a cheaper and feasible alternative for analyzing bone health. The CBA has a sensitivity of 1.0 and a specificity of 0.83, with an area under the Receiver Operating Characteristic curve of 0.93. Our goal was to utilize existing data from primary literature sources to determine if the CBA could be applied with similar or equal fidelity.

Methods: We obtained mean values from analyses of serum Osteocalcin (s-OC) and serum Pyridinoline (s-PYD) markers in conjunction with patient age from various large-sample data sets available in primary literature.

Results: Following analyses of aggregated mean values from the literature, we found that 60% of studies predicted the presence or absence of osteoporosis with the same degree of accuracy between FRAX and CBA methods. Osteoporosis was defined as having a t-score of <−2.5 (FRAX) or surpassing the threshold p-value of >0.035 (CBA).

Conclusions: We expected higher agreement between the FRAX scores and our CBA, but this may be due to the aggregated nature of the data. Our findings indicated the need to advance the CBA in analyzing larger-scale primary data sets, underscoring the importance of raw data analysis, to determine the full efficacy of this diagnostic tool.

Keywords: Chapman bone algorithm, Osteocalcin, Pyridinoline, FRAX

INTRODUCTION

Osteoporosis is an increasingly prevalent metabolic bone disorder affecting over 200 million people worldwide.¹ Characterized by a severe weakening of the bone, osteoporosis is the main cause in the number of hip fractures found in the elderly population annually (approximately 1.6 million cases), of which the number is predicted to reach between 4.5 million and 6.3 million by the year 2050.² While the current “gold standard” model for diagnosing osteoporosis is through determinations of bone mineral density (BMD) from dual-energy x-ray absorptiometry (DXA) technology, recent research has given hope for another, less expensive, and more practical method for determining patient bone health through biomarker “screening” in blood draws.³ The Chapman Bone Algorithm (CBA) utilizes values of patient biomarkers serum Osteocalcin (s-OC) and serum Pyridinoline (s-PYD), in conjunction with patient age to establish an area under the receiver operating characteristic curve as well as sensitivity and 1-specificity correlation coefficients for DXA scan values to determine an individual’s bone health. Equations 1 and 2 reference the components of the CBA. The CBA’s...
accuracy in a relatively small sample of subjects yielded a 93% effectiveness value in conjunction with previously-obtained DXA scores for patients that received blood draws.³

Bone is a metabolically active tissue that undergoes continuous remodeling throughout an individual’s life by the means of two separate physiological processes: mineralization and degradation.⁴ Osteoblasts are specialized bone cells that arise from mesenchymal stem cell progenitors and align themselves in a single layer along advancing edge bone surfaces or previously-osteoclastic tunnels to secrete bone matrix, which is known as bone mineralization.⁵ Osteocalcin (OC) is a 5.8 kDa hydroxyapatite-binding protein synthesized by osteoblasts that expresses calcium-binding properties.⁶ At the carboxy terminus, OC interacts with cell surface receptors and is predisposed as the active molecule in the organization of extracellular matrix.⁷ In contrast, osteoclasts are large, multinucleated bodies that send villus-like projections toward the bone, creating a rough border, and the villi secrete proteolytic enzymes to degrade “old” bone.⁷ Osteoclasts, derived from hematopoietic cells, reabsorb bone through a specialized mechanism⁸. Osteoclastic activity is stimulated by parathyroid hormone (PTH); PTH receptors lie on osteoblasts, and, when stimulated, activate the receptor activator of NF-κB ligand (RANKL) to bind to the receptor activator of NF-κB (RANK) receptors on osteoclasts.⁸ When RANKL binds to RANK receptors, osteoclastogenesis ensues, promoting the degradation of bone. To inhibit osteoclastic activity, osteoprotegerin (OPG) can bind in RANKL’s place on the RANK receptor, allowing for bone formation/generation.⁸ Pyridinoline (PYD) is formed in the extracellular maturation of fibrillar collagens and breaks down cross-linked collagenous during the bone resorption process.⁶ Values of PYD can be obtained from either urine or blood serum. Maintaining the balance between bone formation and degradation is essential in managing healthy bone density and skeletal integrity; those who cannot attain this balance and whose metabolic activity favors osteoclastic activity are prone to osteopenic and osteoporotic conditions.⁹ Because metabolic bone turnover markers (BMs), such as osteocalcin and pyridinoline, represent the activity of bone generation and degradation, it is thought that they would be a feasible method for obtaining info relative to bone health via blood draw.

DXA scanning is the gold standard for determining a patient’s BMD in modern medicine, but is a relatively young technological practice, only dating back to the mid-1980’s. Since the late 1800’s, when skeletal radiography was used to determine patient bone health (requiring roughly 30% bone loss before osteoporosis could be detected via x-rays), bone imaging has progressed tremendously, such that anyone who utilizes DXA technology can determine BMD as normal, osteoporotic, or osteopenic.¹⁰ While the advances in technology have provided more accurate and timely diagnoses for patient bone health, the accessibility for this technology largely remains the paramount criticism. In addition, DXA maintenance can be expensive and requires expert-level knowledge for operation. Due to the economic hardships faced globally by underserved communities and developing nations, the need for establishing alternative methods for diagnostic treatment of metabolic diseases, such as osteoporosis, is pressing.

DXA scans allow healthcare professionals to determine individuals’ bone health in measures of mass per area (g/cm²) from T-scores and Z-scores acquired via a specific mathematical algorithm.¹¹ Using statistical measures of variance from age-adjusted means, individuals can determine their bone mineral density as normal, below-normal (a condition known as osteopenia), or significantly below-normal (indicative of osteoporosis).¹² While DXA has been utilized as the “gold standard” for years as the most accurate way in diagnosing osteoporosis and determining bone health, scans are expensive and difficult to obtain in countries outside the United States, especially in low-income areas or developing nations. To put it into perspective, new DXA scanners equipped with the World Health Organization Fracture Risk Assessment (FRAX),¹³ calculation tool cost more than $100,000 per machine to purchase, leaving many countries unable to provide its services for the impoverished population.¹⁴ Conversely, a simple blood draw, centrifuged locally and mailed to a lab for testing, would be feasible and economically viable. Recent data concerning DXA scanner distribution illustrates the discrepancy in coverage per capita for countries worldwide.¹⁵

While the original CBA study was conducted on a small group of Orange County subjects, the purpose of the current study was to determine if the CBA could be applied with similar or equal fidelity on existing aggregated data sets from the primary literature in the field.¹⁶ This cost-effective and efficient method can be applied in developing nations where current osteoporotic diagnostic tests are unavailable or too expensive.

**METHODS**

**Subject enrollment**

The previous study had a sample size of 226 individuals with information necessary for studying bone health with the CBA.³ Statistical analyses require larger sample sizes to denote important findings relative to the population. For further analysis of the relationship between subjects’ bone biomarkers/age and bone mineral density (BMD), the CBA was conducted on aggregated data sets across 28 different sample groups (aggregate data from 987 individuals in total).

**Procedure and tests performed**

Aggregated data was collected from database publications. Qualifying data sets included the variables...
of interest: values for s-OC, s-PYD, subject age, and BMD.

The primary purpose of the study was to determine the fidelity of the CBA, given a threshold sensitivity and 1-specificity of 0.035 (Equation 2), established from previous research. CBA p-value calculations of subject information of 0.035 or greater exhibited the indication of osteoporosis, whereas values of less than 0.035 did not exhibit the presence of osteoporosis. CBA p-values (Equation 1) were compared with T-scores given subject BMD values (determined by FRAX index), such that p>0.035 and FRAX T-scores < -2.5 were indicative of osteoporosis. If both values predicted or did not predict osteoporosis for a given data set, then CBA and FRAX predictions were in agreement. The relationship between CBA p-values and associated FRAX T-scores were utilized in the original CBA research study, which noted a high accuracy, such that the area under the Receiver Operating Characteristic (ROC) curve was 0.93 in determining the reliability of the algorithm for correctly predicting the incidence of osteoporosis in conjunction with BMD T-scores from DXA scans.

Equation 1: The CBA’s p value determination given variables of interest

\[ \hat{P}(Y = 1 | OC, PYD, Age) = \frac{e^{9.20 + 0.399 OC - 0.459 PYD + 0.089 Age}}{1 + e^{9.20 + 0.399 OC - 0.459 PYD + 0.089 Age}} \]

Equation 2: This model predicted log-odds of having osteoporosis and are given by

\[ \hat{Y} | OC, PYD, Age = \begin{cases} 0, & P(Y = 1 | OC, PYD, Age) \leq 0.035, \\ 1, & P(Y = 1 | OC, PYD, Age) > 0.035. \end{cases} \]

DXA scans for subjects included analysis of (a) lumbar, (b) distal radius, (c) calcaneus, (d) hip, (e) femur, and (f) nonspecific full body average regions. Subject information of BMD from DXA reports were utilized to determine aggregated T-Score values. By convention, T-scores ranging from -1.0 and greater are considered normal for bone density, whereas scores between -1.0 and -2.5 are indicative of an osteopenic state, and scores ranging from -2.5 and lower are indicative of osteoporosis. Aggregated data for s-OC and s-PYD had to be adjusted to fit CBA, such that measurements of both BTMs were adjusted to nanomoles per liter (nmol/L). We used the molecular weight for Osteocalcin as 5,900 g mol⁻¹, and for Pyridinoline we used 428.4 g mol⁻₁ to adjust the diverse ways in which different labs presented the data these metabolites.

RESULTS

We conducted logistic regression modeling consistent with the CBA study for the respective aggregated data sets. Using Equation 1, CBA p-values were determined for aggregate samples using values for mean subject age, s-OC, and s-PYD levels. FRAX T-score values were determined using subject BMD means and standard deviations. With knowledge of previously calculated threshold values for osteoporosis from CBA (Equation 2) and FRAX index, the agreement of CBA and FRAX scores for samples was analyzed. If the threshold for osteoporosis was met or surpassed for both equations for a given study or if both thresholds were not met, then the two algorithms were said to agree with one another for the given sample. As well, the two algorithms were said to not agree with one another for a given sample if thresholds were not met/surpassed or failed to be met. Of the 10 primary literature studies (consisting of 28 different sample groups), 60.7% of the aggregated data sets yielded results with FRAX T-scores and CBA p-values in agreement with one another. The remaining 39.2% of data sets did not produce FRAX T-score and CBA p-values in agreement with one another.

Table 1: Study groups and associated aggregated values found from the literature. Calculations of serum biomarkers s-OC and s-PYD used with subject age to determine CBA p-value; BMD used to determine FRAX T-scores.

<table>
<thead>
<tr>
<th>Study number/Group</th>
<th>CBA p value</th>
<th>Proj. FRAX T-score</th>
<th>CBA/FRAX agreement?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1, Group A¹⁷</td>
<td>4.77E-02</td>
<td>-2.98</td>
<td>Yes</td>
</tr>
<tr>
<td>Study 1, Group B¹⁷</td>
<td>4.77E-02</td>
<td>-4.46</td>
<td>Yes</td>
</tr>
<tr>
<td>Study 2, Group A¹⁹</td>
<td>4.06E-12</td>
<td>-4.46</td>
<td>No</td>
</tr>
<tr>
<td>Study 2, Group B¹⁹</td>
<td>4.06E-12</td>
<td>-4.48</td>
<td>No</td>
</tr>
<tr>
<td>Study 3, Baseline¹⁷</td>
<td>6.65E-05</td>
<td>1.41</td>
<td>Yes</td>
</tr>
<tr>
<td>Study 3, Final¹⁷</td>
<td>2.06E-02</td>
<td>1.32</td>
<td>Yes</td>
</tr>
<tr>
<td>Study 4ª</td>
<td>3.37E-06</td>
<td>-0.90</td>
<td>Yes</td>
</tr>
<tr>
<td>Study 5, Baseline¹⁹</td>
<td>4.21E-12</td>
<td>-4.43</td>
<td>No</td>
</tr>
<tr>
<td>Study 5, Final¹⁹</td>
<td>4.21E-12</td>
<td>-4.35</td>
<td>No</td>
</tr>
<tr>
<td>Study 6⁸</td>
<td>1.60E-02</td>
<td>-12.02</td>
<td>No</td>
</tr>
<tr>
<td>Study 7, Group A¹⁴</td>
<td>7.91E-09</td>
<td>1.99</td>
<td>Yes</td>
</tr>
<tr>
<td>Study 7, Group B¹⁴</td>
<td>7.91E-09</td>
<td>0.60</td>
<td>Yes</td>
</tr>
<tr>
<td>Study 7, Group C¹⁴</td>
<td>1.67E-08</td>
<td>1.64</td>
<td>Yes</td>
</tr>
<tr>
<td>Study 8, Group D¹²</td>
<td>1.67E-08</td>
<td>0.63</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Continued.
DISCUSSION

Earlier research determined that the CBA can be used as a means to detect osteoporosis.\(^3\) In this analysis of aggregated data, we found a 61% agreement between the FRAX index and the CBA in determining the presence of osteoporosis. Our findings suggest that the CBA’s capacity may be limited when dealing with aggregated datasets. However, we believe that the value of the CBA is not diminished by the equivocal findings here; instead, the emphasis should be placed on the need to analyze primary subject data for age, s-OC, and s-PYD. Therefore, we maintain that the CBA may have utility as a cost-effective means of detection in rural areas or developing nations.

As stated previously, the area under the receiver operating characteristic curve in the original study was 0.93 for sensitivity and 1-specificity. When looking into statistical error, neither error type is preferred, but Type II error is generally more tolerated. Therefore, when analyzing the results from this study, a better indication of whether or not the CBA is accurate with FRAX, one should analyze the error presented by the CBA for cases that don’t agree. Of the 10 cases with observed difference between CBA p-value threshold and FRAX T-score threshold values, 90% represent false negatives (Type II error), leaving the remaining 10% as a false positive (Type I error). Thus, even though the CBA p-values do not correctly coincide with the FRAX values, the error nature is more tolerated than that of the false positive category.

Apart from the CBA, other methods have been constructed to analyze bone health and the detection of osteoporosis, despite the ideality of DXA scanning. The trabecular bone score (TBS) is an exemplary alternative to DXA scanning, as TBS takes measurements of the microarchitecture of bones, determining information about bone health independent of BMD and FRAX, most notably, bone propensity to fracture.\(^26\) When added to the presence of common clinical risk factors for osteoporosis (i.e. low calcium intake, benzodiazepine use, height loss), TBS showed an enhanced discriminatory power in predicting a major clinical fracture.\(^27\) Additional methods for self-modulated analysis of bone health have been constructed, such as the Osteoporosis Self-Assessment Tool for Asians (OSTA) and the International Osteoporosis Foundation’s (IOF) tests.\(^28\) OSTA and IOR gave low-, medium-, and high-risk grades for metabolic bone diseases, determining a subject’s relative need to seek a physician’s assistance.

Part of the significance in the development of these alternative methods for detecting early onset osteoporosis lies within the fact that the global elderly population, which exhibits a high prevalence of metabolic diseases like osteoporosis, is estimated to double or triple within the next 30 years.\(^25\) An increasing elderly population would suggest a higher incidence of osteoporosis and bone fractures, of which there are global discrepancies. In 2004, samples of various global populations detailed the prevalence of vertebral bone fractures, such that more than 20% of the populations of elderly women in Japan, North America, Scandinavia, and the Middle East, among others, experienced morphometric vertebral fractures.\(^30\)

Other studies have shown success in using algorithmic approaches for diagnostic purposes in medicine.\(^31,33\) A study centered around the diagnosis of osteoporotic vertebral compression fracture (OVCF) used an algorithmic approach.\(^31\) While slightly different from the necessary components needed for patient determination of osteoporosis with the CBA, the variables for OVCF fracture included patient age, presence/absence of leg pain, BMI, exercise frequency, and gender. Impressively, the algorithm maintained a high AUC (0.95) for sensitivity in conjunction with DXA values for patients; this statistic is important to note because a very large AUC for values in algorithmic prediction models (even in small samples) reflect the diagnostic potential and accuracy of the alternate method(s). On a slightly different note, novel algorithmic diagnostic potential was sought in analyzing patient serum biomarker levels for the initial and follow-up studies on the early detection of osteoarthritis and rheumatoid arthritis, while drawing from similar sample sizes as the CBA (N1=134, N2=225).\(^32,33\) Due to the lack of a blood-based test for predicting a major clinical fracture.\(^27\) Additional methods for self-modulated analysis of bone health have been constructed, such as the Osteoporosis Self-Assessment Tool for Asians (OSTA) and the International Osteoporosis Foundation’s (IOF) tests.\(^28\) OSTA and IOR gave low-, medium-, and high-risk grades for metabolic bone diseases, determining a subject’s relative need to seek a physician’s assistance.

### Table: CBA vs. FRAX Agreement

<table>
<thead>
<tr>
<th>Study number/Group</th>
<th>CBA p value</th>
<th>Proj. FRAX T-score</th>
<th>CBA/FRAX agreement?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 9, Group A (Baseline)(^23)</td>
<td>3.12E-03</td>
<td>0.72</td>
<td>Yes</td>
</tr>
<tr>
<td>Study 9, Group A (Final)(^23)</td>
<td>3.12E-03</td>
<td>0.43</td>
<td>Yes</td>
</tr>
<tr>
<td>Study 9, Group B (Baseline)(^23)</td>
<td>3.12E-03</td>
<td>-0.54</td>
<td>Yes</td>
</tr>
<tr>
<td>Study 9, Group B (Final)(^23)</td>
<td>3.12E-03</td>
<td>-1.17</td>
<td>No</td>
</tr>
<tr>
<td>Study 9, Group C (Baseline)(^23)</td>
<td>2.96E-03</td>
<td>0.42</td>
<td>Yes</td>
</tr>
<tr>
<td>Study 9, Group C (Final)(^23)</td>
<td>2.96E-03</td>
<td>0.13</td>
<td>Yes</td>
</tr>
<tr>
<td>Study 9, Group D (Baseline)(^23)</td>
<td>2.96E-03</td>
<td>-0.85</td>
<td>Yes</td>
</tr>
<tr>
<td>Study 9, Group D (Final)(^23)</td>
<td>2.96E-03</td>
<td>-1.28</td>
<td>Yes</td>
</tr>
<tr>
<td>Study 10, Group A (Baseline)(^24)</td>
<td>8.49E-04</td>
<td>-6.50</td>
<td>No</td>
</tr>
<tr>
<td>Study 10, Group A (Final)(^24)</td>
<td>1.43E-03</td>
<td>-51.75</td>
<td>No</td>
</tr>
<tr>
<td>Study 10, Group B (Baseline)(^24)</td>
<td>4.68E-04</td>
<td>0.30</td>
<td>Yes</td>
</tr>
<tr>
<td>Study 10, Group B (Final)(^24)</td>
<td>1.37E-03</td>
<td>0.50</td>
<td>Yes</td>
</tr>
</tbody>
</table>
osteoarthritis and a lackluster diagnostic model for detection of rheumatoid arthritis, Ahmed, et al, sought out to utilize a cheaper solution and reached modest success in obtaining a diagnostic algorithm with high sensitivity and 1-specificity for the classifications: osteoarthritis, rheumatoid arthritis, and non-rheumatoid arthritis. Parallels can be drawn between these studies and the original CBA study, demonstrating the potential for modern, cheaper serum-based biomarker models that serve as alternative strategies for medical diagnosis in areas where expensive diagnostic machinery may not be an option.

The major implications surrounding the CBA are on the topic of the algorithm’s quick and cost-effective nature for analyzing bone health, suggesting its potential benefit to areas experiencing health disparities due to low socioeconomic conditions and insufficient resources.

The underlying concern for the lack of sufficient osteoporosis detection goes beyond bone health in underserved populations, as osteoporosis has shown comorbid association with other chronic health conditions, including ischemic cardiovascular disease and HIV.

ACKNOWLEDGEMENTS

This study was enabled by a generous grant from the Irvine Health Foundation, which was instrumental for the original study.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: Not required

REFERENCES
