ORIGINAL ARTICLE

Simulating Dual-Energy X-Ray Absorptiometry in CT Using Deep-Learning Segmentation Cascade

SA-CME

Arun Krishnaraj, MD, MPH[a], *Spencer Barrett, MD*[a], *Orna Bregman-Amitai, MSc*[b], *Michael Cohen-Sfady, PhD*[b], *Amir Bar, MSc*[b], *David Chettrit, MSc*[b], *Mila Orlovsky, MSc*[b], *Eldad Elnekave, MD*[b]

Credits awarded for this enduring activity are designated "SA-CME" by the American Board of Radiology (ABR) and qualify toward fulfilling requirements for Maintenance of Certification (MOC) Part II: Lifelong Learning and Self-assessment. To access the SA-CME activity visit https://cortex.acr.org/Presenters/CaseScript/CaseView?CDId=qFix8sPStKA%3d.

Abstract

Purpose: Osteoporosis is an underdiagnosed condition despite effective screening modalities. Dual-energy x-ray absorptiometry (DEXA) screening, although recommended in clinical guidelines, remains markedly underutilized. In contrast to DEXA, CT utilization is high and presents a valuable data source for opportunistic osteoporosis screening. The purpose of this study was to describe a method to simulate lumbar DEXA scores from routinely acquired CT studies using a machine-learning algorithm.

Methods: Between January 2010 and September 2014, 610 CT studies of the abdomen and pelvis were used to develop spinal column and L1 to L4 multiclass segmentation. DEXA simulation training and validation used 1,843 pairs of CT studies accompanied by DEXA results obtained within a 6-month interval from the same individual. Machine learning–based regression was used to determine correlation between calculated grade (on the basis of vertebrae L1-L4) and DEXA *t* score.

Results: Analysis of the *t* score equivalent, generated by the algorithm, revealed true positives in 1,144 patients, false positives in 92 patients, true negatives in 245 patients, and false negatives in 212 patients, resulting in an accuracy of 82%. Sensitivity for the detection of osteoporosis or osteopenia was 84.4% (95% confidence interval, 82.3%-86.2%), and specificity was 72.7% (95% confidence interval, 67.7%-77.2%).

Conclusions: The presented algorithm can identify osteoporosis and osteopenia with a high degree of accuracy (82%) and a small proportion of false positives. Efforts to cull greater information using machine-learning algorithms from pre-existing data have the potential to have a marked impact on population health efforts such as bone mineral density screening for osteoporosis, in which gaps in screening currently exist.

Key Words: DEXA, machine learning, population health, osteoporosis, segmentation cascade

J Am Coll Radiol 2019;■:■-■. Copyright © 2019 American College of Radiology

INTRODUCTION

Osteoporosis remains a prevalent, burdensome, and markedly underdiagnosed condition despite the availability of effective screening modalities [1-3]. Annually, 2 million fractures are attributed to osteoporosis, resulting in more than 432,000 hospital admissions, nearly 2.5 million medical office visits, and approximately 180,000 nursing home admissions in the United States [4]. Estimated direct costs due to osteoporotic fractures total more than $17 billion [5]. Moreover, fractures of the hip pose a major public health burden for the elderly, because these fractures are a major contributor to

[a]Department of Radiology and Medical Imaging, University of Virginia, Charlottesville, Virginia.

[b]Zebra Medical Vision, Shfayim, Israel.

Corresponding author and reprints: Arun Krishnaraj, MD, MPH, Department of Radiology and Medical Imaging, University of Virginia, 1215 Lee Street, PO Box 800170, Charlottesville, VA 22908; e-mail: arunk@virginia.edu.

Dr Chettrit reports personal fees and stock options, outside the submitted work. Dr Orlovsky reports personal fees and stock options outside the submitted work. Dr Bregman-Amitai has a patent US20160015347A1 issued that is broadly relevant to this work. Dr Cohen-Sfady and Dr Elnekave are employees and stockholders of Zebra Medical Vision. The algorhythmic method described is patented by Zebra. The other authors state that they have no conflict of interest related to the material discussed in this article.

1546-1440/19/$36.00 ■ https://doi.org/10.1016/j.jacr.2019.02.033

morbidity, impairment, decreased quality of life, and mortality in men and women [6].

Early detection and appropriate prophylactic treatment are the cornerstones of management of this public health challenge. Clinical practice guidelines recommend dual-energy x-ray absorptiometry (DEXA) bone mineral density (BMD) screening for all women and men older than 65 years [7,8]. However, DEXA screening remains markedly underutilized. Only 19% to 37% of eligible Medicare beneficiaries undergo BMD testing, and fewer than 20% of all persons with major osteoporosis-related fractures have undergone BMD testing and subsequent pharmacologic intervention [9-13].

In contrast to the underutilization of DEXA, CT utilization is high and consistently rising, with more than 70 million examinations performed annually in the United States alone, including 148.1 studies of the abdomen and pelvis performed per 1,000 Medicare beneficiaries each year [14,15]. Routine CT imaging of the abdomen has been recognized as valuable data source for opportunistic osteoporosis screening [16].

The purpose of our study was to describe a method to simulate lumbar DEXA BMD scores from routinely acquired CT studies of the chest and/or abdomen using a machine-learning algorithm. We report the accuracy of algorithmic BMD scoring compared with that obtained in the same population by DEXA scans performed within an interval of 6 months.

METHODS

Data

All imaging examinations were performed as part of routine clinical practice between January 2010 and September 2014 at more than 10 hospitals within an integrated health care system and collected retrospectively. All personal health information was removed before data acquisition, in compliance with HIPAA standards. The study population comprised 70.8% women and 29.2% men. All patients were 50 to 80 years of age.

Six hundred ten CT studies of the abdomen and pelvis were used in the development of spinal column and L1 to L4 multiclass segmentation. In total, 52% of studies used intravenous contrast.

The distribution of CT protocols is depicted in Table 1.

DEXA simulation training and validation involved 1,843 pairs of CT studies accompanied by DEXA results obtained within a 6-month interval from the same individual. Axial DICOM reconstructions of up to 3-mm slice thickness were used for all further planar reconstructions; raw sinusoidal data were not used.

Methods

A schematic view of the deep-learning framework is provided in Figure 1. The framework combines two key components: vertebrae multiclass segmentation and per-vertebra regression, as described below.

Segmentation Process

The vertebral cortical circumference was extracted and mapped into two dimensions [17]. Of note, vertebra-specific segmentation in this model relies on the presence of intervertebral discs and may be limited in circumstances of advanced discogenic disease. To enhance vertebral segmentation in such circumstances, Forstner interest points were extracted on the entire image, followed by foreground and background per point classification using a random forest classifier [18].

Coronal and coronal maximum-intensity projection (MIP) reconstructions were generated from axially acquired CT data. The first lumbar vertebral body (L1) was designated as the first non-rib-bearing vertebra identified on a virtual coronal MIP reconstruction. A virtual sagittal reconstruction was then generated, whereby any scoliosis is corrected in relation to a center line on a coronal reconstruction. Multiclass segmentation was then performed, and on the basis of the vertebral segmentation, a simulated DEXA score could be computed for L1 to L4 vertebrae.

The multiclass segmentation approach is based on a cascade of two U-Nets (Figure 2). First, binary segmentation on the vertebral column is performed in a sagittal view. Then multiclass segmentation is performed on the basis of coregistration of each

Table 1. Distribution of CT protocols

Study Type	% of Total
CT abdomen pelvis	61
CT abdomen and chest	31
CT chest	5
CT urography	2
CT skeletal	1
CT colonography	1

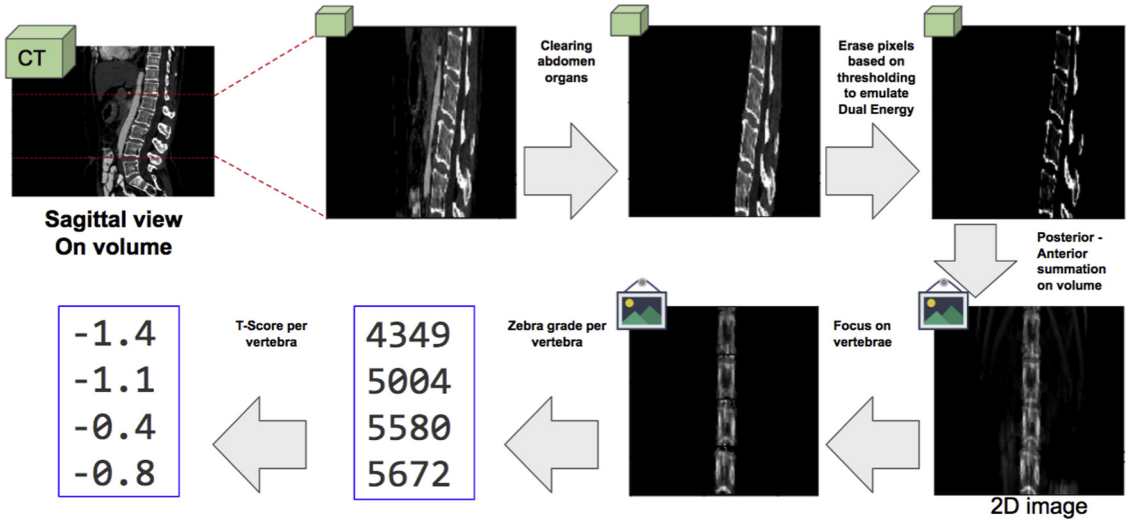

Fig 1. Schematic view of the deep-learning framework.

vertebral body from the virtual sagittal and coronal MIP reconstructions. Applying multiclass segmentation solely on the basis of the sagittal view resulted in poor performance because of lack of anatomic context provided by the ribs. The segmentation process is visually described in Figure 3.

Unlike patch-based segmentation, fully convolutional neural networks add up-sampling layers to standard convolutional neural networks to recover the spatial resolution of the input at the output layer. To compensate for the resolution loss induced by pooling layers, fully convolutional neural networks introduce skip connections between their down-sampling and up-sampling paths.

DEXA Simulation

The volume anterior and posterior to the vertebral column was removed from the 3-D image. To identify the optimal pixel range intensity for *t* score simulation, three machine-learning methodologies were used: linear regression, support vector machine, and logistic regression models. Linear regression delivered the best performance. The specific weights per vertebra (L1-L4) were as follows: 0.00116455, 0.00106969, 0.00110523, and 0.00078648. The specific biases per vertebra were −6.43475268, −6.69634826, −6.65447089, and −5.39662082, respectively.

Machine learning–based regression methodologies were thus used to determine the range of pixel intensities in which the simulated *t* score and the DEXA *t* score were best correlated. X-ray anteroposterior acquisition was simulated, ignoring pixels outside the learned intensity range to provide a summation map. The generated information was used to evaluate the simulated *t* score on the basis of vertebrae L1-L4.

The average virtual *t* score for L1 to L4 showed little variation on average (Table 2). Simulated *t* scores were transformed to a categorical results at a threshold value of −1.

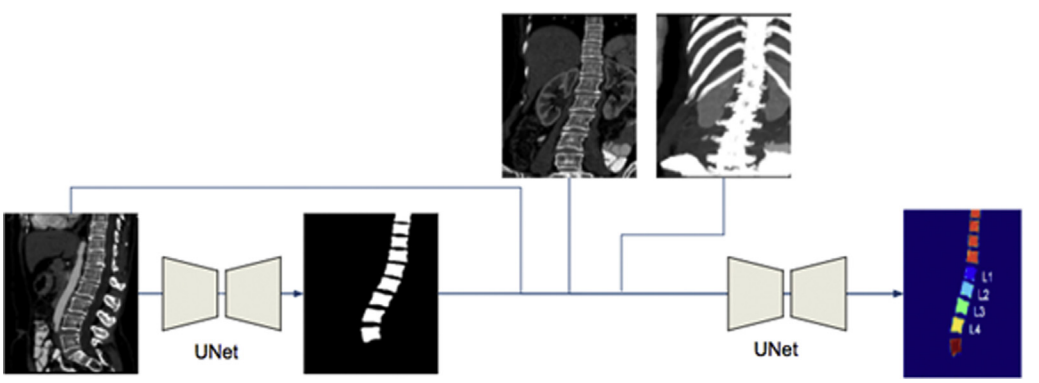

Fig 2. Vertebrae lumbar multiclass segmentation framework.

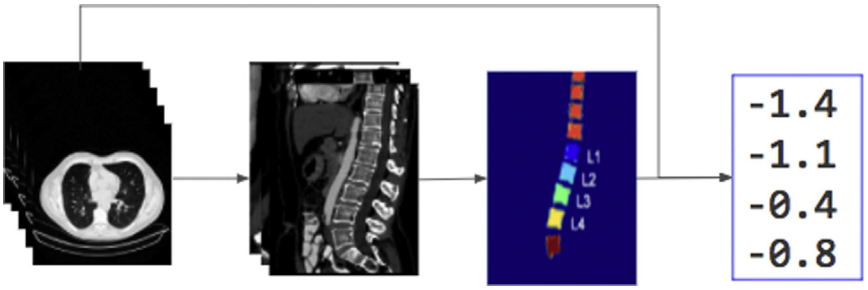

Fig 3. Visual description of the segmentation process using CT data. Key views of the patient vertebrae are extracted. Multiclass segmentation is then performed, and on the basis of the vertebral segmentation, a simulated dual-energy x-ray absorptiometry score is computed for L1 to L4 vertebrae.

For each obtained *t* score equivalent value, it was determined whether this value was

- true positive (TP): abnormal/abnormal (if the *t* score equivalent value is <-1 and the DEXA scan result is abnormal [ie, <-1]);
- false positive (FP): abnormal/normal (if the *t* score equivalent value is <-1 and the DEXA scan result is normal [ie, ≥ -1]);
- true negative (TN): normal/normal (if the *t* score equivalent value is ≥ -1 and the DEXA scan result is normal [ie, ≥ -1]); or
- false negative (FN): normal/abnormal (if the *t* score equivalent value is ≥ -1 and the DEXA scan result is abnormal [ie, <-1]).

The categorical dichotomous result of normal was returned for all *t* score equivalent scores ≥ -1; the categorical dichotomous result of abnormal was returned for all *t* score equivalent scores < -1.

RESULTS

Achievement of the *t* Score Equivalent Value

A total of 1,693 CT studies with corresponding DEXA imaging within a 6-month interval were used for validation testing of the algorithm. The segmentation algorithm was trained, tested, and validated on a separate data set of more than 10,000 CT studies. The CT-to-DEXA conversion algorithm was fit on the basis of 20% of the 1,693 CT-DEXA combinations. These were included in the validation analysis of the entire set of 1,693 CT-DEXA combinations, representing a risk for overfitting.

Table 2. Average virtual *t* scores for L1 to L4

	L1	L2	L3	L4
Average virtual *t* score	−1.47	−1.59	−1.03	−1.16
SD	1.44	1.58	1.65	1.67

t Score Equivalent Compared With Recorded DEXA *t* Score

Analysis of the *t* score equivalent provided by the presented algorithm revealed TPs for 1,144 patients, FPs for 92 patients, TNs for 245 patients, and FNs for 212 patients.

Specificity, Sensitivity, and Accuracy of the Software

In total, results for 1,389 patients were defined as "true" (ie, concordant with the recorded DEXA results), resulting in an accuracy of 82%. The software detected osteoporosis or osteopenia in the vast majority of cases with calculated sensitivity of 84.4% (95% confidence interval, 82.3%-86.2%). The specificity was 72.7% (95% confidence interval, 67.7%-77.2%). The Pearson correlation achieved was 0.8524 (Fig. 4); the Bland-Altman results were $1.96 \times SD = 1.56$ (Fig. 5). Of note, no significant difference between the algorithmic predictive accuracy was noted when testing men and women separately.

DISCUSSION

Osteoporosis affects one-third of women and one-fifth of men older than 50 years. The prevalence of osteoporosis is expected to increase as the global population ages because of the inverse relationship between bone density and age. Moreover, 80% of those at risk are not identified or treated. Approximately 2 million fractures occur annually in the United States because of the consequences of unaddressed osteoporosis, resulting in approximately $17 billion in health care expenditures. Patients who sustain osteoporotic fractures experience significant degradation in their quality of life: 25% of patients with hip fractures end up in nursing homes within 12 months

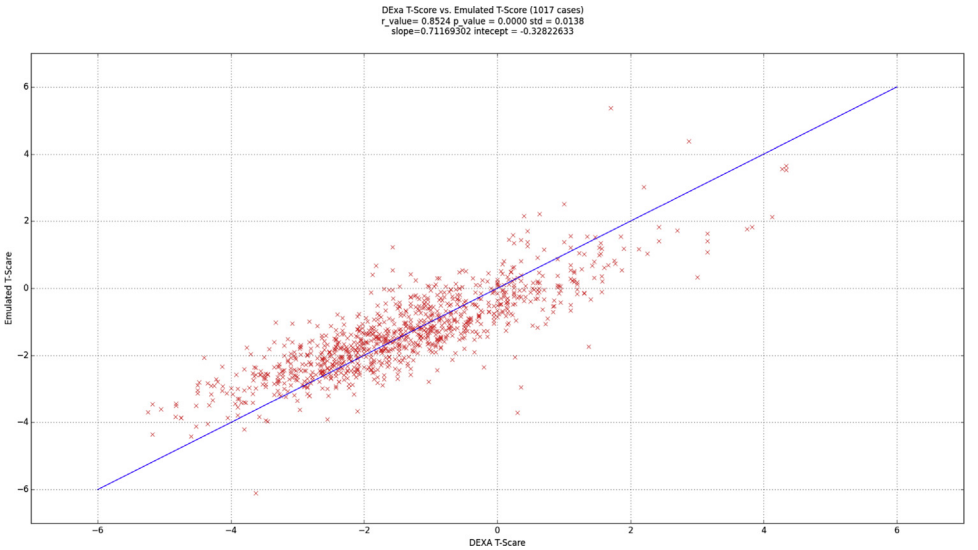

Fig 4. Pearson correlation: dual-energy x-ray absorptiometry (DEXA) *t* score versus simulated *t* score.

of their fractures. Despite the detrimental effects and the availability of effective prophylactic treatments, a very small proportion of the population older than 50 years actively undergoes DEXA scans, the gold standard for diagnosis of osteoporosis.

CT scans are performed much more often than DEXA scans in patients older than 50 years and hence provide an opportunity to identify those at risk for osteoporotic fracture. Abdominal and chest CT scans are obtained for various clinical indications. The interpreting radiologist may comment on particular destructive or blastic bone lesions, but it is uncommon for a radiologist to comment on overall BMD or to suggest a diagnosis of osteoporosis on the basis of CT data. Routine manual or visual assessment of bone mineralization on every CT scan of the chest and abdomen would significantly alter radiologist workflow and decrease work productivity. Even osteoporotic vertebral body fractures seen on CT are unlikely to be reported, unless the clinical indication for the study relates directly to that observation [19].

In our analysis, using chest and abdominal CT scans from more than 1,000 patients, we demonstrate that the presented algorithm can identify osteoporosis and osteopenia with a high degree of accuracy (82%) and a small proportion of FPs.

On the basis of these results, the majority of patients with underlying osteoporosis or osteopenia who undergo chest or abdominal CT can be accurately identified using our proposed CT-targeted DEXA

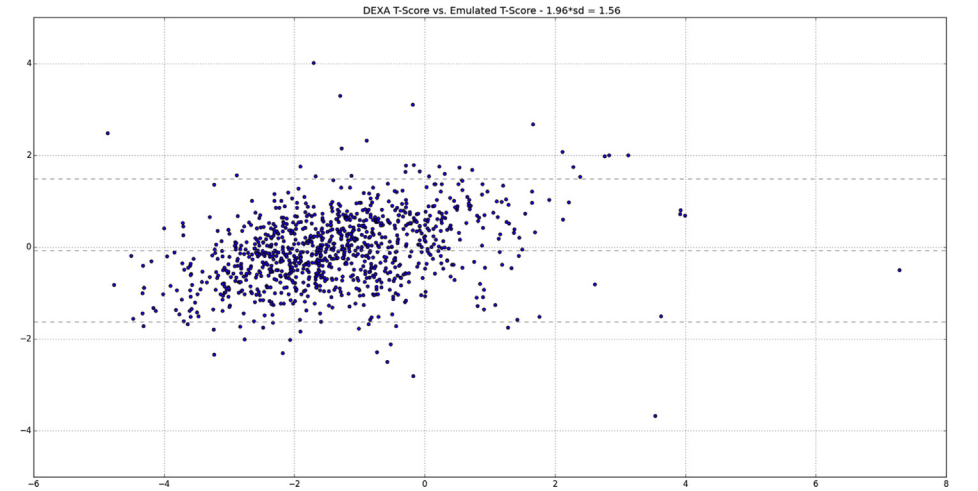

Fig 5. Bland-Altman results: dual-energy x-ray absorptiometry (DEXA) *t* score versus simulated *t* score.

screening. To our knowledge, previous similar-sized studies have not produced CT *t* scores using a fully automated tool that allows large-scale implementation and could not reach a similar accuracy.

On the basis of extrapolation of our health maintenance organization's volume of DEXA and CT scans, a CT-targeted DEXA screening initiative has the potential to increase annual osteoporosis diagnoses by approximately 50%, with a 2.7-fold improvement in pretest probability.

Our study had the following potential limitations: restricting the source area for the simulation to exclusively the lumbar vertebrae, using DEXA as the basis for training, and using data only from a single integrated health system. DEXA scans typically sample multiple sites, including the radius and ulna, femur, and lumbar vertebrae, which are at high risk for fracture, to provide a more sensitive means of detecting osteoporosis, which can be regionally heterogeneous. Restricting the use of CT data to the lumbar vertebrae may reduce sensitivity by failing to detect clinically significant osteoporotic areas outside the lumbar spine. As in other studies that were based on vertebral measures, we did not assess risk factors used in various fracture risk calculators, such as the FRAX tool. However, using data from other high-risk areas such as the femur is inherently more complex and remains an area of future research. Additionally, DEXA scans, which were used as the standard for testing the validity of using CT data to detect osteoporosis in this study, are known to have limitations. Most notably, DEXA has decreased sensitivity for osteoporosis in the case of significant degenerative disease. The potential consequence of this is an erroneously increased FP rate when testing the simulated DEXA against actual DEXA data. Last, there are potential limitations to the generalizability of the study to a broader population. Variation between populations in terms of factors such as age, comorbidities, and health maintenance habits or differences in CT or DEXA technique between health systems may confound the ability to extrapolate the results of this study. Moreover there is no standard interval between CT and DEXA to determine BMD metrics. We chose 6 months to accrue enough retrospective data in a reasonable time frame. We did not accrue enough studies to perform a statistically significant analysis of correlation for the interval between DEXA and CT. We intend to do so in future work.

Efforts to cull greater information using machine-learning algorithms from pre-existing data, as demonstrated in our study, have the potential to have a marked impact on population health efforts such as BMD screening for osteoporosis, in which gaps in screening currently exist.

TAKE-HOME POINTS

- Osteoporosis remains a prevalent, burdensome, and markedly underdiagnosed condition despite the availability of effective screening modalities.
- Although recommended by clinical practice guidelines, the current gold standard for osteoporosis screening, DEXA, remains markedly underutilized.
- In contrast to the underutilization of DEXA, CT utilization is high, presenting a valuable data source for opportunistic osteoporosis screening.
- The machine learning–based algorithms presented in our study can identify osteoporosis and osteopenia from routinely acquired CT data with a high degree of accuracy.
- Efforts to cull greater information using machine-learning algorithms from pre-existing data has the potential to have a marked impact on population health efforts such as BMD screening for osteoporosis, in which gaps in screening currently exist.

REFERENCES

1. Cummings SR, Melton LJ. Epidemiology and outcomes of osteoporotic fractures. Lancet 2002;359:1761-7.
2. Kanis JA, Delmas P, Burckhardt P, Cooper C, Torgerson D. Guidelines for diagnosis and management of osteoporosis. The European Foundation for Osteoporosis and Bone Disease. Osteoporos Int 1997;7:390-406.
3. Kanis JA. Diagnosis of osteoporosis and assessment of fracture risk. Lancet 2002;359:1929-36.
4. (Office of the Surgeon General. Bone health and osteoporosis. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20945569. Accessed December 28, 2017.
5. Burge R, Dawson-Hughes B, Solomon DH, Wong JB, King A, Tosteson A. Incidence and economic burden of osteoporosis-related fractures in the United States 2005-2025. J Bone Miner Res 2007;22:465-75.
6. Pisani P, Renna MD, Conversano F, et al. Major osteoporotic fragility fractures: risk factor updates and societal impact. World J Orthop 2016;7:171-81.
7. Looker AC, Orwoll ES, Johnston CC, et al. Prevalence of low femoral bone density in older U.S. adults from NHANES III. J Bone Miner Res 1997;12:1761-8.
8. Brown JP, Fortier M, Frame H, et al. Canadian Consensus Conference on osteoporosis 2006 update. J Obstet Gynaecol Can 2006;28(2 suppl 1):S95-112.
9. Gillespie CW, Morin PE. Trends and disparities in osteoporosis screening among women in the United States 2008-2014. Am J Med 2017;130:306-16.
10. King AB, Fiorentino DM. Medicare payment cuts for osteoporosis testing reduced use despite tests' benefit in reducing fractures. Health Aff (Millwood) 2011;30:2362-70.

11. Curtis JR, Carbone L, Cheng H, et al. Longitudinal trends in use of bone mass measurement among older Americans 1999-2005. J Bone Miner Res 2008;23:1061-7.
12. McAdam-Marx C, Unni S, Ye X, Nelson S, Nickman NA. Effect of Medicare reimbursement reduction for imaging services on osteoporosis screening rates. J Am Geriatr Soc 2012;60:511-6.
13. Leslie WD, Giangregorio LM, Yogendran M, et al. A population-based analysis of the post-fracture care gap 1996-2008: the situation is not improving. Osteoporos Int 2012;23:1623-9.
14. Berrington de González A, Mahesh M, Kim K-P, et al. Projected cancer risks from computed tomographic scans performed in the United States in 2007. Arch Intern Med 2009;169:2071-7.
15. Levin DC, Rao VM, Parker L, Frangos AJ, Sunshine JH. Bending the curve: the recent marked slowdown in growth of noninvasive diagnostic imaging. AJR Am J Roentgenol 2011;196:W25-9.
16. Pickhardt PJ, Pooler BD, Lauder T, del Rio AM, Bruce RJ, Binkley N. Opportunistic screening for osteoporosis using abdominal computed tomography scans obtained for other indications. Ann Intern Med 2013;158:588.
17. Yao J, Burns JE, Forsberg D, et al. A multi-center milestone study of clinical vertebral CT segmentation. Comput Med Imaging Graph 2016;49:16-28.
18. Kulkarni R, Tuller M, Fink W, Wildenschild D. Three-dimensional multiphase segmentation of x-ray CT data of porous materials using a Bayesian Markov random field framework. Vadose Zone J 2012;11.
19. Carberry GA, Pooler BD, Binkley N, Lauder TB, Bruce RJ, Pickhardt PJ. Unreported vertebral body compression fractures at abdominal multidetector CT. Radiology 2013;268:120-6.