



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/167712/>

Version: Published Version

Article:

Almog, Y.A., Rai, A., Zhang, P. et al. (2020) Deep learning with electronic health records for short-term fracture risk identification : crystal bone algorithm development and validation. *Journal of Medical Internet Research*, 22 (10). e22550.

<https://doi.org/10.2196/22550>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Original Paper

Deep Learning With Electronic Health Records for Short-Term Fracture Risk Identification: Crystal Bone Algorithm Development and Validation

Yasmeen Adar Almog¹, BSE; Angshu Rai¹, BTECH, MTECH; Patrick Zhang¹, BSc, MS; Amanda Moulaison¹, BSc; Ross Powell¹, BA; Anirban Mishra¹, BSc, MSc; Kerry Weinberg¹, BSc, MBA, MSc; Celeste Hamilton², BHK, MSc, PhD; Mary Oates³, BSc, MD; Eugene McCloskey⁴, MD, FRCP; Steven R Cummings⁵, BA, MD, FACP

¹Digital Health & Innovation, Amgen Inc, Thousand Oaks, CA, United States

²Global Medical Operations, Amgen Inc, Thousand Oaks, CA, United States

³US Medical, Amgen Inc, Thousand Oaks, CA, United States

⁴Department of Oncology & Metabolism, The University of Sheffield, Sheffield, United Kingdom

⁵Department of Medicine, University of California San Francisco, San Francisco, CA, United States

Corresponding Author:

Yasmeen Adar Almog, BSE
Digital Health & Innovation
Amgen Inc
1 Amgen Center Drive
MS 38-3B
Thousand Oaks, CA, 91320
United States
Phone: 1 4243463036
Email: yalmog@amgen.com

Abstract

Background: Fractures as a result of osteoporosis and low bone mass are common and give rise to significant clinical, personal, and economic burden. Even after a fracture occurs, high fracture risk remains widely underdiagnosed and undertreated. Common fracture risk assessment tools utilize a subset of clinical risk factors for prediction, and often require manual data entry. Furthermore, these tools predict risk over the long term and do not explicitly provide short-term risk estimates necessary to identify patients likely to experience a fracture in the next 1-2 years.

Objective: The goal of this study was to develop and evaluate an algorithm for the identification of patients at risk of fracture in a subsequent 1- to 2-year period. In order to address the aforementioned limitations of current prediction tools, this approach focused on a short-term timeframe, automated data entry, and the use of longitudinal data to inform the predictions.

Methods: Using retrospective electronic health record data from over 1,000,000 patients, we developed Crystal Bone, an algorithm that applies machine learning techniques from natural language processing to the temporal nature of patient histories to generate short-term fracture risk predictions. Similar to how language models predict the next word in a given sentence or the topic of a document, Crystal Bone predicts whether a patient's future trajectory might contain a fracture event, or whether the signature of the patient's journey is similar to that of a typical future fracture patient. A holdout set with 192,590 patients was used to validate accuracy. Experimental baseline models and human-level performance were used for comparison.

Results: The model accurately predicted 1- to 2-year fracture risk for patients aged over 50 years (area under the receiver operating characteristics curve [AUROC] 0.81). These algorithms outperformed the experimental baselines (AUROC 0.67) and showed meaningful improvements when compared to retrospective approximation of human-level performance by correctly identifying 9649 of 13,765 (70%) at-risk patients who did not receive any preventative bone-health-related medical interventions from their physicians.

Conclusions: These findings indicate that it is possible to use a patient's unique medical history as it changes over time to predict the risk of short-term fracture. Validating and applying such a tool within the health care system could enable automated and widespread prediction of this risk and may help with identification of patients at very high risk of fracture.

KEYWORDS

fracture; bone; osteoporosis; low bone mass; prediction; natural language processing; NLP; machine learning; deep learning; artificial intelligence; AI; electronic health record; EHR

Introduction

Fractures due to osteoporosis and low bone mass are associated with a significant personal, clinical, and economic burden. These fractures are common; the risk of sustaining such a fracture increases with age, and their incidence is expected to increase worldwide as the population ages [1-11]. In the United States, an estimated 1 in 2 women and 1 in 4 men over 50 years of age will experience such a fracture [12-14]. However, there remains a significant diagnosis and treatment gap for osteoporosis [1,2,4,12]. When these fractures occur, they often result in a loss of independence for patients and can lead to functional disability, lower quality of life, and increased mortality [5,15-38]. Given this substantial burden and unmet need for interventions, it is critical to identify patients at risk of fracture, as effective management of risk can prevent these deleterious outcomes.

Several fracture risk prediction tools have been developed for clinical use. The most commonly used tools are the University of Sheffield Fracture Risk Assessment Tool, known as FRAX [39], and the Garvan Institute of Health Bone Fracture Risk Calculator (GIH-BFRC) [40]. Both tools use a set of cross-sectional clinical risk factors to evaluate fracture likelihood, and typically require manual data entry to perform the predictions. The performance of both methods varies greatly in real-world analyses; this variance is partially explained by study population and design and predicted fracture outcome (hip vs other osteoporotic fractures). In a review [41], 12 studies of FRAX showed an average area under the receiver operating characteristics curve (AUROC) of 0.65 (SD 0.038) when predicting major osteoporotic fractures without including bone mineral density in the model, and similar results were shown for GIH-BFRC [41]. These commonly used risk assessment tools estimate 5- and 10-year fracture risk but do not provide estimates of 1- to 2-year risk [42-45].

Increased risk of fracture in the next 1-2 years is not routinely assessed in clinical practice, despite the existence of rapid-acting preventative therapeutics [8,46,47]. Although methods for predicting short-term risk have been explored [48-50], they have not yet been widely clinically accepted. Furthermore, these models are limited to a specific set of cross-sectional information, some of which may not readily be available. Thus, there remains a need to further develop a fracture risk prediction tool that predicts on a short-term time frame in order to facilitate identification of patients at high risk. While there are published examples [51-53] applying artificial intelligence to fracture and osteoporosis risk, these approaches focus either on imaging data [51] or on cross-sectional data for long-term predictions [52,53]. To our knowledge, there is no existing method that applies deep learning to sequential patient data for predicting fracture risk.

To address these unmet needs, we developed *Crystal Bone*, a machine learning approach that leverages techniques typically applied in natural language processing. However, rather than applying these methods to text-based data, we applied them to longitudinal data contained in electronic health records. Specifically, we focused on diagnosis codes (International Classification of Diseases; ICD), treating each code as a word and sequences of codes as stories. The goal of this study was to evaluate the ability of these natural language processing-based models to learn patterns associated with increased short-term (ie, 2-year) fracture risk. The results of our analyses suggest that not only does this unique longitudinal method produce accurate short-term fracture risk predictions, but also that the approach can help fulfill the unmet need that exists in fracture-risk identification.

Methods

Data Background

We used subsets of the Optum deidentified electronic health record data set, which contains comprehensive longitudinal electronic health record data for 91 million patients from over 140,000 providers (as of March 2018) from the United States. The subsets, which contain bone health and pan-therapeutic populations respectively, cover the time from January 1, 2007, through December 31, 2018 (Optum, email communication, August 2019).

The bone health subset was obtained by filtering for patients with osteoporosis, fractures, or bone-related medications (n=6,329,986). In the period covered by the data set, the fracture incidence rate (ie, the proportion of fractures among all events detected, which may include multiple fractures per person) was 39% in the population over 50 years of age. The bone health data set was primarily used for training the model.

The pan-therapeutic data set represented a random sample of 5% of the overall Optum electronic health record data set and contained patient data (n=3,476,219) with no filtering for any specific comorbidities or treatments; this dataset had a fracture incidence rate of 8.5% in the population over 50 years of age. Because the sample was drawn from such a large population, the pan-therapeutic data set was assumed to be broadly representative of the US population. As such, we performed all model evaluations on a testing sample from this data set (a holdout data set), to better understand the generalizability of the model in a real-world setting.

Ethical Approval

Since this was a retrospective study using deidentified data, patients were not required to actively participate in the study. Therefore, neither informed consent of patients nor institutional review board approval was required.

Data Engineering and Cohort Selection

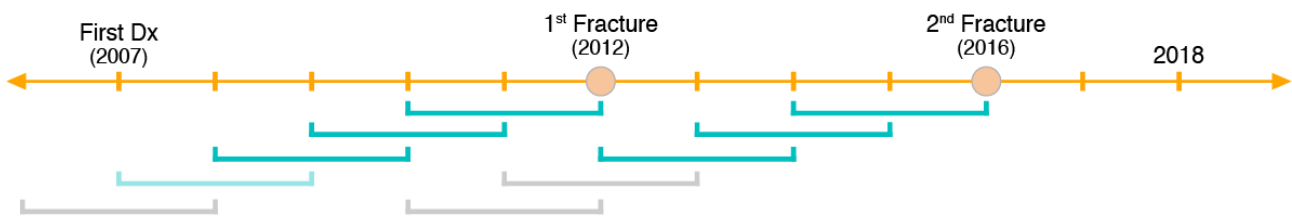
The cohort consisted of patients who were at least 50 years of age at the time of their event; this criterion was chosen to reduce the data to a population that is more susceptible to fractures associated with osteoporosis and low bone mass. For fracture patients, an event is the date of occurrence of any qualifying fracture. Qualifying fractures are defined by a set of rules based on those used by Wright et al [54] for identifying novel and relevant fracture events in claims data. For nonfracture patients, an event is the date of the last recorded diagnosis of any kind in the data set. We describe further details of the fracture identification process in Multimedia Appendix 1.

We further filtered our cohorts for patients with at least 2 years of medical history leading up to their respective events. Applying these parameters limited the bone health cohort to 3,408,494 patients and the pan-therapeutic cohort to 700,315 patients.

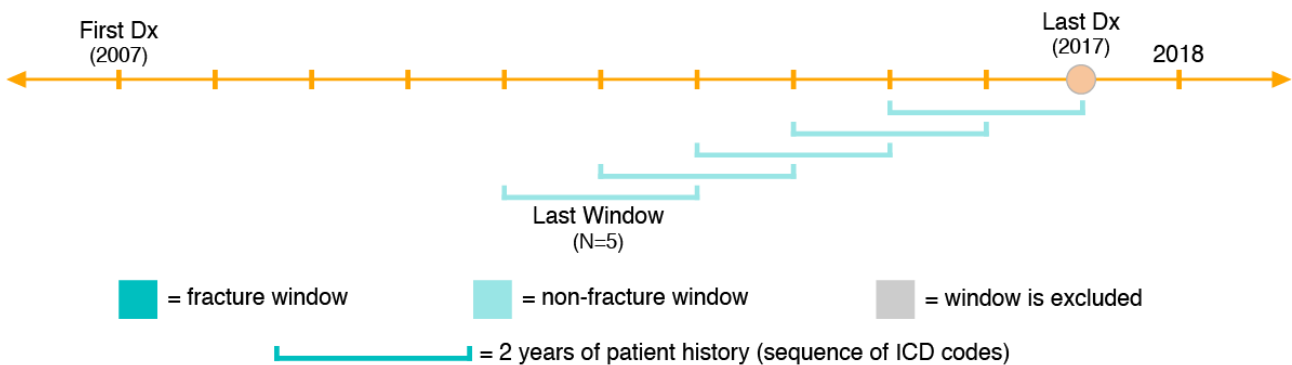
We applied sliding windows to the data (Figure 1), where each event could have up to 5 windows, and each window was a historical sequence defined as the list of chronologically ordered ICD codes in the 2 years leading up to an event. These historical sequences were then used to predict risk of fracture within a 2-year horizon (a 1-year horizon was also explored, see Multimedia Appendix 1). As shown in Figure 1, some windows were dropped from the analysis due to incomplete or potentially overlapping coverage. Additionally, windows that occur more than 2 years before a fracture event were labeled as nonfracture windows. The motivation for this approach was to provide the algorithm with multiple unique code sequences leading up to the same event that may reflect changes in risk at various times within the given time horizon. Furthermore, the fixed window size provided a consistent timeframe for prediction as opposed to varying lengths of time for each patient, which would have occurred if patients' complete code histories were used. Further details regarding the motivation and methodology of this approach are in Multimedia Appendix 1.

Figure 1. Sliding window algorithm schematic. This schematic depicts the sliding window algorithm for a multifracture and nonfracture patient. Dx:diagnosis; ICD: International Classification of Diseases.

Sample Fracture Patient Windows:



Sample Non-Fracture Patient Windows:



There was no additional filtering based on specific diagnoses or comorbidities. For each qualifying patient, the algorithms utilized all available ICD codes in the historical sequences described above. Only the codes that occurred fewer than 5 times in the full cohort were excluded, as these codes were too rare to be included in the diagnosis code vocabulary.

Data Sampling

Before model training, we generated a 70:30 random split of the pan-therapeutic data, representing training and holdout subsets. Since the pan-therapeutic data set is highly imbalanced, with a fracture event incidence of only 6.5% after applying the sliding window algorithm, we oversampled additional fracture windows from the bone health data set to achieve a balanced

(50:50) training set for modeling. This oversampling training paradigm was replicated for all models. The holdout set remained untouched, with the original distribution of fractures.

Modeling Approaches

Overview

Crystal Bone was inspired by techniques that are typically applied in natural language processing. However, instead of applying these techniques to text-based data, we applied them to sequences of ICD codes. Correspondingly, each ICD code was analogous to a word, and each sequence of ICD codes was analogous to a document. To this end, we implemented 2 distinct frameworks: (1) ICD code vectorization and long short-term

memory networks, and (2) patient-level vectorization and extreme gradient boosting decision trees. Both approaches utilize sequences of ICD codes as inputs. The ICD code vectorization and long short-term memory framework undertakes this task by first learning semantic definitions for the codes, then evaluating the sequence of definitions through a deep learning network. The patient-level vectorization and extreme gradient boosting modeling framework employs a similar approach; however, rather than embedding individual ICD codes, it embeds the entire ICD code sequence for each patient, thereby learning “summaries” of patient sequences. This framework produces a prediction by feeding these summaries through a decision tree classifier. The model parameters were tuned to optimize AUROC; details of this process are provided in [Multimedia Appendix 1](#).

Framework 1: ICD Code Vectorization + Long Short-Term Memory

The first framework consisted of 2 primary components. The ICD code vectorization component was responsible for learning a “definition” for each ICD code based on skip-gram architecture word embedding (word2vec) [55], an unsupervised learning approach that mapped each code in the vocabulary to a 100-dimensional vector. To generate these embeddings, we utilized sequences from the pan-therapeutic training set alone (without oversampling), to avoid bias toward bone-health related codes. In our implementation, the vocabulary consisted of all diagnosis codes that occurred at least 5 times in this data set, amounting to more than 40,000 unique codes. The method generated a vector for each code based on the context in which it appeared; in electronic health records, similar ICD codes

appear in similar contexts, and as a result have similar vector representations. These embeddings reduced the dimensionality and sparsity of the feature space, and helped the neural network recognize related ICD codes. [Figure 2](#) illustrates the encoded vectors projected onto a 2D space using uniform manifold approximation and projection (UMAP) for dimension reduction [56]. The collocation of related diagnosis codes in this coordinate space provided qualitative evidence that the ICD code vectorization had encoded meaningful latent information.

The long short-term memory component consisted of a neural network with long short-term memory layers, a deep learning architecture that enables the evaluation of recurrent data, such as sequences of embedded ICD codes. We trained this network with the complete training set (including oversampling from the bone health data set). The long short-term memory network predicted the likelihood of a fracture event within 2 years as a classification problem. Long short-term memory networks are a common approach for solving such problems [57].

Additionally, given the ubiquitous use of nonsequential features such as age and sex for predicting fracture risk, we supplied age and sex to the neural network as static features through concatenation of long short-term memory and dense layers. Furthermore, because the long short-term memory framework required all input sequences to have uniform length, we also included total diagnosis count as a static feature to account for the effects of truncating or padding the sequences. The schematic in [Figure 3](#) provides an overview of the model architecture and inputs to the algorithm, namely age, sex, diagnosis count, and the patient’s unique sequence of ICD codes.

Figure 2. 2D projection of ICD-10 code embeddings from the ICD code vectorization model: (a) All ICD-10 codes by the first letter (high-level category) of the code, (b) a cluster of codes related to alcohol near coordinates (2.3, 3) by code subgroups, (c) a cluster of codes related to kidney function near coordinates (3.75, 0.025) by code subgroups, and all ICD-10 fracture codes in region C (d) by region of the body, and (e) by frequency of occurrence. ICD: International Classification of Diseases; UMAP: uniform manifold approximation and projection.

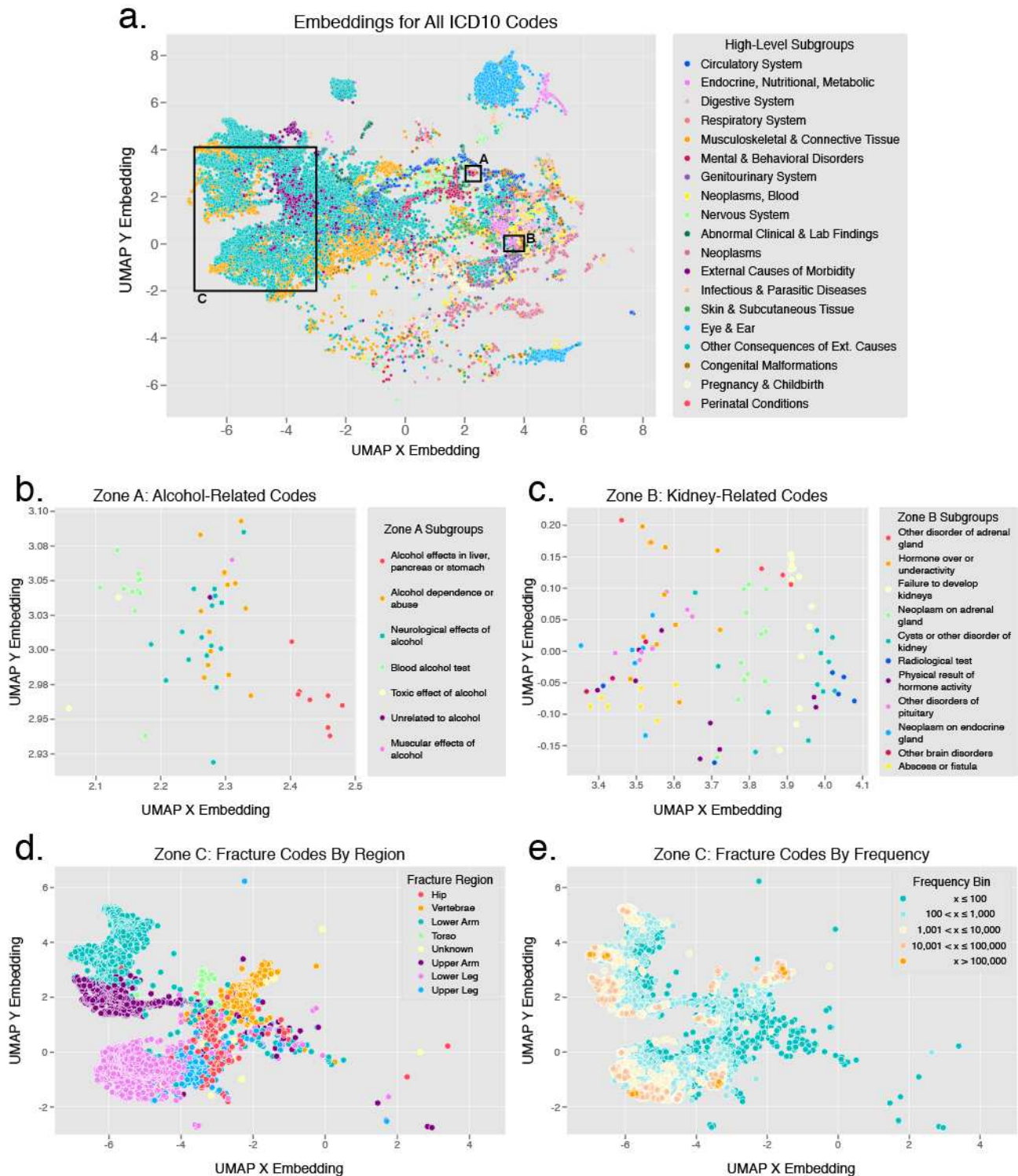
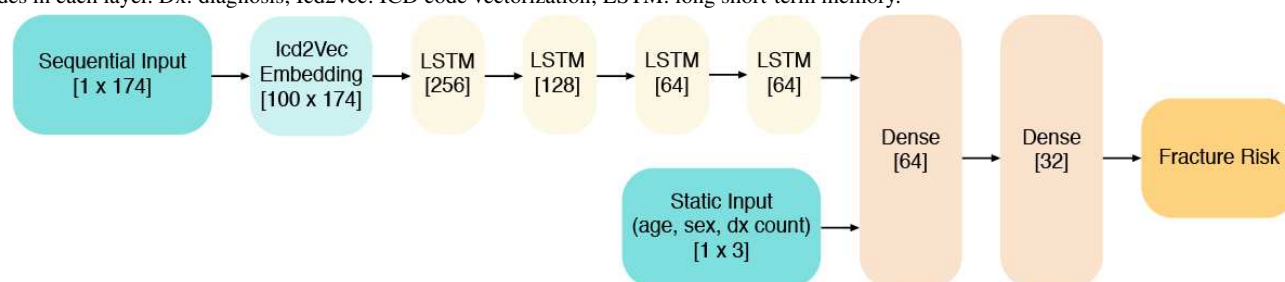


Figure 3. High-level architecture of the long short-term memory neural network including the dimensionality of the inputs, as well as the number of nodes in each layer. Dx: diagnosis; Icd2vec: ICD code vectorization; LSTM: long short-term memory.



Framework 2: Patient-Level Vectorization and Extreme Gradient Boosting

Similar to the ICD code vectorization + long short-term memory modeling framework, the patient-level vectorization and extreme gradient boosting decision trees framework consists of 2 components. First, the patient-level vectorization embeds entire ICD code sequences to a 128-dimensional semantic space using the distributed bag of words framework [58]. Much as the ICD code vectorization learned definitions of individual ICD codes, the patient-level vectorization instead learned summaries of patient sequences. The method for doing so is the same; patients with similar sequential contexts will have similar summary vectors. We trained the patient-level vectorization with the sliding window ICD code sequences, again only utilizing the pan-therapeutic data to avoid bias toward the bone health therapeutic area. This created embeddings that represented 2-year episodes of patient histories; a detailed exploration of these embeddings is in [Multimedia Appendix 1](#).

The extreme gradient boosting decision trees component utilized the embeddings from the patient-level vectorization, as well as the static features of age, sex, and total diagnosis count that were incorporated in Framework 1, to predict fracture risk. This type of algorithm, also referred to as XGBoost, is a scalable tree-based modeling approach that improves the generalizability, speed, and efficacy of prediction [59]. We trained this algorithm with the full training set (including bone health data set oversampling) to learn a classification model that predicted the likelihood of fracture within 2 years.

Ensemble Model

An ensemble model was also evaluated. This algorithm combined the outputs of both the aforementioned frameworks with a logistic regression metaclassifier.

Baseline Models

We compared these modeling frameworks to 2 baseline models. The first baseline model utilized the age and sex of each patient at each window. These were 2 of the only features shared by the FRAX tool and the GIH-BFRC models. The other shared feature is prior fracture; however, because neither the FRAX tool nor GIH-BFRC's method of measuring this value was possible for our data set without censoring, we did not include it in the model. The second baseline incorporated age, sex, and total diagnosis count (number of ICD codes) in each sample; these represent all of the static features used by both modelling frameworks, enabling evaluation of the relative benefit of including sequential ICD code data. Both baseline models utilized extreme gradient boosting decision tree algorithms, the same classification approach that was used in Framework 2.

Human-Level Performance Approximation

In addition to these baselines, we approximated human-level performance by isolating a set of retrospective physician-prescribed interventions that were identifiable in the electronic health record data set. These interventions consisted of diagnostic tests as well as pharmacologic treatments. The list of interventions was based on treatment guidelines provided by the National Osteoporosis Foundation [60] and the Journal of Clinical Endocrinology and Metabolism [61] and was further validated by the physician coauthors of this manuscript, who confirmed that the interventions aligned with their understanding of osteoporosis treatment guidelines ([Table 1](#)). If a patient received one of these interventions in a 2-year historical window, that window was flagged as "physician-identified risk, worthy of intervention." A full description of the limitations of this approach is described in [Multimedia Appendix 1](#).

Table 1. List of physician interventions for human-level performance analysis.

Type and name	Pharmacologic
Procedure	
Dual-energy x-ray absorptiometry	No
Vertebral fracture assessment	No
Quantitative computed tomography	No
Other bone density measurements (single energy x-ray absorptiometry, radiographic absorptiometry, ultrasound, single-photon absorptiometry)	No
Bone turnover markers	No
Administration of any medications referenced below	Yes
Treatment	
Bisphosphonates (alendronate, alendronate-cholecalciferol, ibandronate, risedronate, zoledronic acid)	Yes
Abaloparatide	Yes
Denosumab	Yes
Raloxifene	Yes
Bazedoxifene	Yes
Romozosumab	Yes
Teriparatide	Yes
Calcitonin	Yes
Diagnosis	
Osteoporosis (M80, M81, 733.0)	No

We defined the cohort of patients who did not receive any form of intervention (diagnoses, tests, or treatments) as *no intervention* and assessed how well the algorithm was able to correctly identify which patients had a fracture within 2 years, as well as how frequently the algorithm mistakenly flagged patients with no imminent fracture. We also evaluated the patients who received interventions (the *intervention cohort*) with this method, referred to as the *cohort analysis*. However, since an intervention can directly modulate fracture risk, we performed a separate analysis in order to mitigate some of the uncertainty due to the effects of interventions. For this analysis, we identified each patient's first pharmacologic intervention and used the diagnosis history leading up to this date as input. This analysis allowed us to gauge the extent to which the algorithm flags agreed with human-level performance interventions (without needing to adjust for their effects). We termed this the *overlap analysis*. The cohort analysis utilized

the full list of interventions, while the overlap analysis utilized the pharmacological subset of the list of interventions.

Model Performance

We report model performance on a set of 5 primary metrics: AUROC, recall (sensitivity), specificity, precision, and area under the precision-recall curve (AUPRC).

Results

Model Performance

The overall performance of the algorithms is shown through comparison of the 2 frameworks with the 2 baseline models to demonstrate the quality of each algorithm's predictions. [Table 2](#) shows a summary of key model performance metrics on the same holdout data set. The Crystal Bone models, including the ensemble model that combined the 2 approaches, outperformed the baseline models for nearly all performance metrics.

Table 2. Comparison of model performance metrics.

Model	AUROC ^a	Recall	Specificity	Precision	AUPRC ^b
ICD code vectorization + LSTM ^c	0.812	0.646	0.812	0.192	0.462
Patient level vectorization + XGBoost ^d	0.790	0.670	0.758	0.161	0.358
Ensemble	0.818	0.693	0.777	0.177	0.463
Baseline (age, sex)	0.667	0.787	0.416	0.0855	0.119
Baseline (age, sex, diagnosis count)	0.668	0.547	0.707	0.114	0.130

^aAUROC: area under the receiver operating characteristics curve.

^bAUPRC: area under the precision-recall curve.

^cLSTM: long short-term memory.

^dXGBoost: extreme gradient boosting.

ICD Code Vectorization + Long Short-Term Memory Model

To further characterize this performance, we evaluated the ICD code vectorization and long short-term memory model on primary and subsequent fracture events. While the model performs best on subsequent fractures, both primary and subsequent fracture analyses (AUROC 0.742 and 0.910, respectively) show a marked improvement against corresponding baseline models (AUROC 0.591 and 0.747, respectively). We report detailed results of this experiment and additional evaluations of sensitivity and robustness of this model in [Multimedia Appendix 1](#).

Human-Level Performance Comparison

[Table 3](#) contains the results of the cohort analysis. For windows with no interventions, Crystal Bone Framework 1 correctly flagged 16,127 of the 28,626 windows that resulted in fracture

(56.3%); this corresponds to 9649 out of 13,765 (70.1%) of the unique fracture events. Crystal Bone Framework 1 incorrectly flagged 91,717 of the 532,621 windows with no fractures as at-risk (17.2%); however, 1053 of the windows in this cohort (3%) sustained a fracture in >2 years.

For windows with interventions, only 11,833 of 69,198 (17.1%) of the detected interventions included treatments; thus, the remaining 57,365 (82.9%) interventions were either diagnoses or diagnostic tests. In the intervention cohort, Crystal Bone Framework 1 correctly captured 10,277 out of 12,244 windows for which fracture occurred within 2 years (83.9%). For the windows with interventions and no fracture event, 19,235 out of 56,954 (33.8%) are incorrectly flagged by our algorithm as at risk. These results suggest Crystal Bone's ability to recognize interventions through their associated ICD codes and adjust the predicted fracture risk accordingly. However, a deeper exploration of specific interventions is required to verify this.

Table 3. Human-level performance results.

Cohort	Windows, n (%)	Flag, n (%)	No flag, n (%)
Total	630,445 (100)	— ^a	—
No intervention	561,247 (89.0)	—	—
Fracture	28,626 (5.1)	16,127 (56.3)	12,449 (43.7)
Nonfracture	532,621 (94.9)	91,717 (17.2)	440,904 (82.8)
Intervention	69,198 (11.0)	—	—
Fracture	12,244 (17.7)	10,277 (83.9)	1967 (16.1)
Nonfracture	56,954 (82.3)	19,235 (33.8)	37,719 (66.2)

^aNot reported.

The overlap analysis enabled us to better understand how well Crystal Bone Framework 1 correlated with observed physician interventions through exploration of the first pharmacological treatment in the holdout set. Of the 7127 patients who received treatment, 6071 had enough medical history leading up to this treatment for Crystal Bone Framework 1. When evaluating these patients, 3017 out of those 6071 (49.7%) were considered at risk of fracture in 2 years.

We evaluated the incidence of fracture within 2 years for this subgroup. Of the cohort deemed at risk by the algorithm, 684

out of 3017 (22.7%) experienced a fracture within 2 years of the first intervention date. This precision is a slight improvement over that of the algorithm on the overall holdout set, at 19.2%. Furthermore, of all 570 patients in this pharmacological intervention cohort who ultimately suffered from a fracture within 2 years, Crystal Bone Framework 1 correctly flagged 469 (82.3%).

Discussion

General

In this study, we evaluated the performance of 2 natural language processing–inspired fracture prediction models: (1) ICD code vectorization and long short-term memory (AUROC 0.812) and (2) patient-level vectorization and extreme gradient boosting (AUROC 0.790). The performance of these models reflected a substantial improvement over 2 baseline models: (1) with age and sex (AUROC 0.670) and (2) with age, sex, total diagnosis count (AUROC 0.670). Furthermore, these short-term prediction metrics were an improvement over cross-sectional tools for long-term time frames, such as FRAX and GIH-BFRC, which have been widely clinically accepted [41]. Although fundamental differences in study design make it impossible to compare these metrics directly, sensitivity analyses of Crystal Bone across fracture types, prediction time frames, and fracture definitions suggest robust predictive performance and generalizability. To our knowledge, this is the first study that has experimented with separate models for primary and subsequent fracture types; further discussion of this analysis, as well as the additional sensitivity analyses, is in [Multimedia Appendix 1](#).

The human-level performance comparison provides deeper insight to the benefits of Crystal Bone. The retrospective labeling utilized in both the cohort and overlap analyses enabled a scalable, data-driven comparison of physician action and Crystal Bone and avoided bias that may occur through alternative methods of human-level performance evaluation [62]. To our knowledge, this is the first fracture risk prediction study which includes such a human-level performance comparison in the analysis.

Through the cohort analysis we learned that only a small proportion of patients received preventative interventions, including basic diagnostic tests, showcasing the extent of unmet need in the health care system [1,2,4,12]. In the subset of patient windows with no interventions, Crystal Bone was able to flag 70.1% of the unique fracture events. Given the existence of rapid-acting preventative therapeutics [8,46,47], as well as the demonstrated efficacy of bone-forming agents in reduction of 1- to 2-year fracture risk [63-69], these results suggest that, had appropriate preventative measures been taken, the risk of these fractures may have been reduced, thus mitigating a significant burden to both the patient and the health care system.

The findings of the overlap analysis further support the merits of Crystal Bone, through demonstration of alignment with observable interventions made by physicians. Because it is impossible to confirm whether these treatment interventions were taken in response to a perceived short-term risk of fracture, we cannot expect 100% overlap between Crystal Bone and these observed interventions. We saw that Crystal Bone was aligned with these physician interventions 49.7% of the time. While this overlap is not complete, it captured 82.3% of the patients who ultimately experienced a fracture, reflecting the algorithm's increased sensitivity for the cohort deemed at-risk by physicians.

This suggests a meaningful alignment with both physician evaluation and actual observed fracture risk. Ultimately, these human-level performance comparisons, coupled with performance against baseline models and alternative risk prediction methods, suggest that Crystal Bone can fulfill a critical unmet need through identification of patients at high risk of fracture.

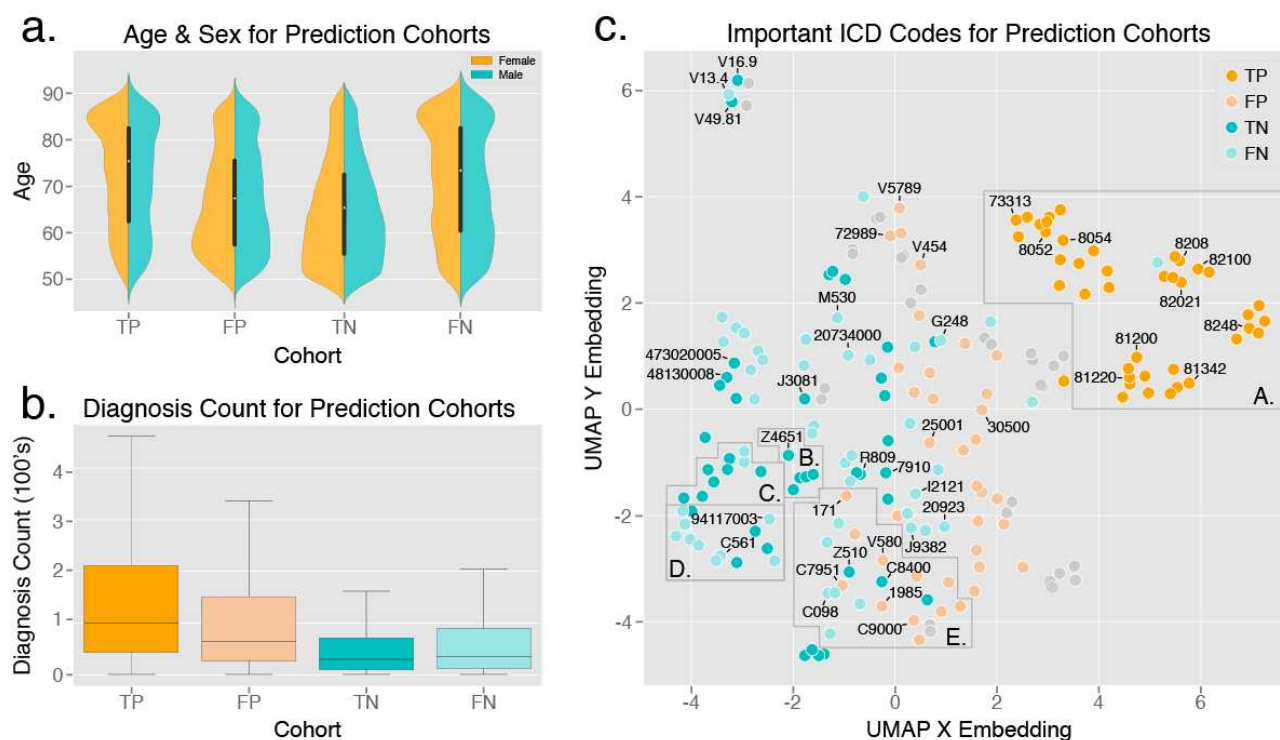
Limitations of the Current Approach

Various limitations exist for the approaches described, particularly from the inherent complications of using real-world data. The techniques described rely upon ICD codes recorded in electronic health record systems, which will impact the performance and validity of the models if diagnoses are not detected, incorrectly recorded, or missed due to patient dropout. Indeed, most vertebral fragility fractures are clinically silent and hence not captured in electronic health records [70]. While an approach utilizing only ICD codes is potentially more comprehensive and straightforward for real-world implementation due to the quality of coverage and descriptive nature of diagnosis codes, we may miss salient clinical features captured elsewhere in the electronic health record. For example, there exist ICD codes associated with obesity, osteopenia, and osteoporosis, which represent measurements of BMI and bone mineral density on a categorical level. However, these do not reflect exact clinical measurements; the exclusion of these quantitative measurements may limit the performance and clinical impact of the algorithm. Nevertheless, it may be advantageous to utilize these ICD codes rather than the quantitative measures, as such measures in an electronic health record frequently contain human error and may not always be readily available.

In addition to data set challenges, there exist limitations inherent to assumptions of the modeling approach. The suppositions of constant time between diagnosis codes and uniform sequence length may affect performance. Exploration of more advanced methods that do not require such assumptions could improve the model and is an area of future work.

Perhaps the greatest limitation of the described approaches is that they are generally considered black box approaches and lack significant interpretability. Developing methods for improved interpretation of deep learning models is an active area of research. We have performed an initial exploration of this for the ICD code vectorization and long short-term memory model in [Figure 4](#), which compares various characteristics of the four prediction cohorts of the confusion matrix for the test set (true positive [TP], false positive [FP], true negative [TN], false negative [FN]). Within each of these groups, we performed exploratory analysis on the associated samples for each of the input features in the model: age, sex, total diagnosis count, and ICD codes. Results of this analysis are described in detail in [Multimedia Appendix 1](#). While this serves as an initial evaluation of model interpretability, a deeper exploration of interpretability techniques is an area for future work in these algorithms.

Figure 4. Exploration of model interpretability by comparison of various characteristics of the input data for the 4 prediction cohorts of the confusion matrix. FN: false negative; FP: false positive; ICD: International Classification of Diseases; TN: true negative; TP: true positive; UMAP: uniform manifold approximation and projection.



Another limitation of this study is the inability to perform direct comparisons with established risk calculators such as FRAX. Additionally, this approach has yet to be validated with external data, which is the subject of future work.

Potential Applications

We foresee numerous applications of this work in the health care system, with benefits for patients, providers, and payers alike. For payers, Crystal Bone provides a unique opportunity to explore population health, enabling insurers to identify and address patients in need of evaluation or intervention, and preventing the large expenses associated with fracture events. For providers, direct electronic health record integration would facilitate patient care, and help identify at-risk patients who are not currently identified as such. That being said, effective implementation requires additional understanding on the impact of interventions on short-term fracture risk; while there is evidence to suggest that rapid acting treatments and bone-forming agents can significantly decrease fracture risk on a shortened time frame [8,46,47,63-69], a more detailed exploration of the optimal care pathways for various Crystal Bone risk scores would likely be required to facilitate real-world use of the algorithm.

Crystal Bone addresses the need for an automated and largely physician-independent tool that is effective at predicting

short-term fracture risk. It is the first such approach that takes longitudinal patient trajectories into account, rather than focusing primarily on cross-sectional information, enabling a more personalized assessment of fracture risk. Furthermore, with automated aggregation of patient histories in an electronic health record system, the prediction of fracture risk could be entirely hands-off, without requiring a doctor or patient to manually enter any information into the software. This unique approach may facilitate broader adoption of the algorithm. Still, the lack of clinical guidelines for 1- and 2-year risk may limit adoption in the near future.

Such a tool, if widely applied, could facilitate early patient identification, and help reduce the morbidity and mortality associated with fractures. The retrospective human-level performance comparison suggests that Crystal Bone would identify patients who are currently missed in the health care system, potentially minimizing the burden on patients and the health care system overall. Given the prevalence and anticipated increase of fractures due to osteoporosis and low bone mass as the population ages, as well as the enormous personal, clinical, and economic costs associated with such fractures, Crystal Bone could provide a meaningful positive impact through reduced burden and improved outcomes.

Acknowledgments

This study was funded by Amgen Inc. The costs covered by Amgen Inc were licensing of the Optum data set, access to the computational resources required to develop the model, and compensation for listed Amgen Inc employees. No additional funding was provided for this study.

Thank you to Optum for providing access to and assistance with the data. We would like to additionally thank the following individuals for their guidance and support in conducting this study and creating this manuscript: Inbal Lapid, Tammy Lindberg, Howard Chen, Mandy Suggitt, Lisa Humphries, Marc Doble, Nkem Ogbechie, John Page, Michi He, Akhila Balasubramanian, and Erle Davis.

Authors' Contributions

YA is the first author. Technical conception, design and direction: YA, AR, PZ, and KW. Medical direction and interpretation: CH, MO, EM, and SRC. Data analysis and interpretation: YA, PZ, AWM, RP and AM. Writing of the manuscript: YA, AR, PZ, KW, CH, MO, EM, and SRC. Authors EM and SR contributed equally. All authors contributed to critical revisions of the draft and approved the final manuscript.

Conflicts of Interest

YA, PZ, RP, AM, KW, and MO are employees and stock owners at Amgen Inc, the funders of this study. AR, AWM, and CH are former employees and stock owners at Amgen Inc. EM is a consulting fee recipient, grant recipient, and speaker on behalf of Amgen Inc, as well as a member of the International Osteoporosis Foundation. SRC is a consulting fee recipient and grant recipient from Amgen Inc.

Multimedia Appendix 1

Supplementary Information.

[\[DOC File , 1001 KB-Multimedia Appendix 1\]](#)

References

1. Haczynski J, Jakimiuk A. Vertebral fractures: a hidden problem of osteoporosis. *Med Sci Monit* 2001;7(5):1108-1117. [Medline: [11535963](#)]
2. Svedbom A, Hernlund E, Ivergård M, Compston J, Cooper C, Stenmark J, EU Review Panel of IOF. Osteoporosis in the European Union: a compendium of country-specific reports. *Arch Osteoporos* 2013 Oct 11;8(1-2):137 [[FREE Full text](#)] [doi: [10.1007/s11657-013-0137-0](#)] [Medline: [24113838](#)]
3. Davies KM, Stegman MR, Heaney RP, Recker RR. Prevalence and severity of vertebral fracture: The saunders county bone quality study. *Osteoporosis Int* 1996 Mar;6(2):160-165. [doi: [10.1007/bf01623941](#)]
4. Facts and Statistics 2015. International Osteoporosis Foundation. URL: [https://www.iofbonehealth.org/facts-statistics#:~:text=Osteoporosis%20is%20estimated%20to%20affect,USA%20and%20Japan%20\(1\)](https://www.iofbonehealth.org/facts-statistics#:~:text=Osteoporosis%20is%20estimated%20to%20affect,USA%20and%20Japan%20(1)) [accessed 2020-01-05]
5. Kanis JA, Johnell O, Oden A, Borgstrom F, Zethraeus N, De Laet C, et al. The risk and burden of vertebral fractures in Sweden. *Osteoporos Int* 2004 Jan;15(1):20-26. [doi: [10.1007/s00198-003-1463-7](#)] [Medline: [14593450](#)]
6. Cooper C, Atkinson EJ, O'Fallon WM, Melton JL. Incidence of clinically diagnosed vertebral fractures: a population-based study in Rochester, Minnesota, 1985-1989. *J Bone Miner Res* 1992 Feb;7(2):221-227. [doi: [10.1002/jbmr.5650070214](#)] [Medline: [1570766](#)]
7. 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 Mar;22(3):465-475 [[FREE Full text](#)] [doi: [10.1359/jbmr.061113](#)] [Medline: [17144789](#)]
8. Lötters FJB, van den Bergh JP, de Vries F, Rutten-van Mólken MPMH. Current and future incidence and costs of osteoporosis-related fractures in The Netherlands: combining claims data with BMD measurements. *Calcif Tissue Int* 2016 Mar;98(3):235-243 [[FREE Full text](#)] [doi: [10.1007/s00223-015-0089-z](#)] [Medline: [26746477](#)]
9. Rosengren BE, Karlsson MK. The annual number of hip fractures in Sweden will double from year 2002 to 2050. *Acta Orthopaedica* 2014 Apr 30;85(3):234-237. [doi: [10.3109/17453674.2014.916491](#)]
10. Gullberg B, Johnell O, Kanis J. World-wide projections for hip fracture. *Osteoporos Int* 1997 Sep;7(5):407-413. [doi: [10.1007/pl00004148](#)]
11. Papadimitropoulos EA, Coyte PC, Josse RG, Greenwood CE. Current and projected rates of hip fracture in Canada. *CMAJ* 1997 Nov 15;157(10):1357-1363. [Medline: [9371065](#)]
12. Office of the Surgeon General (US). Bone Health and Osteoporosis: A Report of the Surgeon General. Rockville (MD): Office of the Surgeon General (US) 2004:67-105. [Medline: [20945569](#)]
13. What is osteoporosis and what causes it? National Osteoporosis Foundation. 2016. URL: <https://www.nof.org/patients/what-is-osteoporosis/> [accessed 2020-01-05]
14. Harvey N, Earl S, Cooper C. Epidemiology of osteoporotic fractures. In: Favus MJ, editor. *Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism* 6th ed. Washington, DC: American Society for Bone and Mineral Research; 2006:244-248.

15. Cooper C. The crippling consequences of fractures and their impact on quality of life. *Am J Med* 1997 Aug 18;103(2A):12S-17S. [doi: [10.1016/s0002-9343\(97\)90022-x](https://doi.org/10.1016/s0002-9343(97)90022-x)] [Medline: [9302893](https://pubmed.ncbi.nlm.nih.gov/9302893/)]
16. Hip Fracture Outcomes in People Age 50 and Over-Background Paper: OTA-BP-H- 120. US Congress Office of Technology Assessment. Washington, DC: US Government Printing Office; 1994 Jul. URL: <http://www.princeton.edu/~ota/disk1/1994/9413/9413.PDF> [accessed 2020-10-01]
17. Tajeu GS, Delzell E, Smith W, Arora T, Curtis JR, Saag KG, et al. Death, debility, and destitution following hip fracture. *J Gerontol A Biol Sci Med Sci* 2014 Mar;69(3):346-353 [FREE Full text] [doi: [10.1093/gerona/glt105](https://doi.org/10.1093/gerona/glt105)] [Medline: [23873945](https://pubmed.ncbi.nlm.nih.gov/23873945/)]
18. Sabesan VJ, Valikodath T, Childs A, Sharma VK. Economic and social impact of upper extremity fragility fractures in elderly patients. *Aging Clin Exp Res* 2015 Aug 24;27(4):539-546. [doi: [10.1007/s40520-014-0295-y](https://doi.org/10.1007/s40520-014-0295-y)] [Medline: [25708827](https://pubmed.ncbi.nlm.nih.gov/25708827/)]
19. Ray NF, Chan JK, Thamer M, Melton LJ. Medical expenditures for the treatment of osteoporotic fractures in the United States in 1995: report from the National Osteoporosis Foundation. *J Bone Miner Res* 1997 Jan 01;12(1):24-35 [FREE Full text] [doi: [10.1359/jbmr.1997.12.1.24](https://doi.org/10.1359/jbmr.1997.12.1.24)] [Medline: [9240722](https://pubmed.ncbi.nlm.nih.gov/9240722/)]
20. Hall SE, Williams JA, Senior JA, Goldswain PR, Criddle RA. Hip fracture outcomes: quality of life and functional status in older adults living in the community. *Aust N Z J Med* 2000 Jun;30(3):327-332. [doi: [10.1111/j.1445-5994.2000.tb00833.x](https://doi.org/10.1111/j.1445-5994.2000.tb00833.x)] [Medline: [10914749](https://pubmed.ncbi.nlm.nih.gov/10914749/)]
21. Marottoli RA, Berkman LF, Cooney LM. Decline in physical function following hip fracture. *J Am Geriatr Soc* 1992 Sep 27;40(9):861-866. [doi: [10.1111/j.1532-5415.1992.tb01980.x](https://doi.org/10.1111/j.1532-5415.1992.tb01980.x)] [Medline: [1512379](https://pubmed.ncbi.nlm.nih.gov/1512379/)]
22. Nevitt MC, Ettinger B, Black DM, Stone K, Jamal SA, Ensrud K, et al. The association of radiographically detected vertebral fractures with back pain and function: a prospective study. *Ann Intern Med* 1998 May 15;128(10):793-800. [doi: [10.7326/0003-4819-128-10-199805150-00001](https://doi.org/10.7326/0003-4819-128-10-199805150-00001)] [Medline: [9599190](https://pubmed.ncbi.nlm.nih.gov/9599190/)]
23. Pasco JA, Henry MJ, Korn S, Nicholson GC, Kotowicz MA. Morphometric vertebral fractures of the lower thoracic and lumbar spine, physical function and quality of life in men. *Osteoporos Int* 2009 May 19;20(5):787-792. [doi: [10.1007/s00198-008-0744-6](https://doi.org/10.1007/s00198-008-0744-6)] [Medline: [18802658](https://pubmed.ncbi.nlm.nih.gov/18802658/)]
24. Fischer S, Kapinos KA, Mulcahy A, Pinto L, Hayden O, Barron R. Estimating the long-term functional burden of osteoporosis-related fractures. *Osteoporos Int* 2017 Oct 24;28(10):2843-2851. [doi: [10.1007/s00198-017-4110-4](https://doi.org/10.1007/s00198-017-4110-4)] [Medline: [28647804](https://pubmed.ncbi.nlm.nih.gov/28647804/)]
25. Dyer SM, Crotty M, Fairhall N, Magaziner J, Beaupre LA, Cameron ID, Fragility Fracture Network (FFN) Rehabilitation Research Special Interest Group. A critical review of the long-term disability outcomes following hip fracture. *BMC Geriatr* 2016 Sep 02;16:158 [FREE Full text] [doi: [10.1186/s12877-016-0332-0](https://doi.org/10.1186/s12877-016-0332-0)] [Medline: [27590604](https://pubmed.ncbi.nlm.nih.gov/27590604/)]
26. Abimanyi-Ochom J, Watts JJ, Borgström F, Nicholson GC, Shore-Lorenti C, Stuart AL, et al. Changes in quality of life associated with fragility fractures: Australian arm of the International Cost and Utility Related to Osteoporotic Fractures Study (AusICUROS). *Osteoporos Int* 2015 Jun 20;26(6):1781-1790 [FREE Full text] [doi: [10.1007/s00198-015-3088-z](https://doi.org/10.1007/s00198-015-3088-z)] [Medline: [25792491](https://pubmed.ncbi.nlm.nih.gov/25792491/)]
27. Brennehan SK, Barrett-Connor E, Sajjan S, Markson LE, Siris ES. Impact of recent fracture on health-related quality of life in postmenopausal women. *J Bone Miner Res* 2006 Jun 06;21(6):809-816 [FREE Full text] [doi: [10.1359/jbmr.0603011](https://doi.org/10.1359/jbmr.0603011)] [Medline: [16753011](https://pubmed.ncbi.nlm.nih.gov/16753011/)]
28. Palacios S, Neyro JL, Fernández de Cabo S, Chaves J, Rejas J. Impact of osteoporosis and bone fracture on health-related quality of life in postmenopausal women. *Climacteric* 2014 Feb 30;17(1):60-70. [doi: [10.3109/13697137.2013.808182](https://doi.org/10.3109/13697137.2013.808182)] [Medline: [23710562](https://pubmed.ncbi.nlm.nih.gov/23710562/)]
29. Roux C, Wyman A, Hooven FH, Gehlbach SH, Adachi JD, Chapurlat RD, GLOW investigators. Burden of non-hip, non-vertebral fractures on quality of life in postmenopausal women: the Global Longitudinal study of Osteoporosis in Women (GLOW). *Osteoporos Int* 2012 Dec 8;23(12):2863-2871 [FREE Full text] [doi: [10.1007/s00198-012-1935-8](https://doi.org/10.1007/s00198-012-1935-8)] [Medline: [22398855](https://pubmed.ncbi.nlm.nih.gov/22398855/)]
30. Crans GG, Silverman SL, Genant HK, Glass EV, Krege JH. Association of severe vertebral fractures with reduced quality of life: reduction in the incidence of severe vertebral fractures by teriparatide. *Arthritis Rheum* 2004 Dec;50(12):4028-4034 [FREE Full text] [doi: [10.1002/art.20671](https://doi.org/10.1002/art.20671)] [Medline: [15593198](https://pubmed.ncbi.nlm.nih.gov/15593198/)]
31. Kado DM, Browner WS, Palermo L, Nevitt MC, Genant HK, Cummings SR. Vertebral fractures and mortality in older women: a prospective study. Study of Osteoporotic Fractures Research Group. *Arch Intern Med* 1999 Jun 14;159(11):1215-1220. [doi: [10.1001/archinte.159.11.1215](https://doi.org/10.1001/archinte.159.11.1215)] [Medline: [10371229](https://pubmed.ncbi.nlm.nih.gov/10371229/)]
32. Bentler SE, Liu L, Obrizan M, Cook EA, Wright KB, Geweke JF, et al. The aftermath of hip fracture: discharge placement, functional status change, and mortality. *Am J Epidemiol* 2009 Nov 15;170(10):1290-1299 [FREE Full text] [doi: [10.1093/aje/kwp266](https://doi.org/10.1093/aje/kwp266)] [Medline: [19808632](https://pubmed.ncbi.nlm.nih.gov/19808632/)]
33. Bliuc D, Nguyen ND, Nguyen TV, Eisman JA, Center JR. Compound risk of high mortality following osteoporotic fracture and refracture in elderly women and men. *J Bone Miner Res* 2013 Nov 18;28(11):2317-2324 [FREE Full text] [doi: [10.1002/jbmr.1968](https://doi.org/10.1002/jbmr.1968)] [Medline: [23616397](https://pubmed.ncbi.nlm.nih.gov/23616397/)]
34. Hu F, Jiang C, Shen J, Tang P, Wang Y. Preoperative predictors for mortality following hip fracture surgery: a systematic review and meta-analysis. *Injury* 2012 Jun;43(6):676-685. [doi: [10.1016/j.injury.2011.05.017](https://doi.org/10.1016/j.injury.2011.05.017)] [Medline: [21683355](https://pubmed.ncbi.nlm.nih.gov/21683355/)]
35. Jette AM, Harris BA, Cleary PD, Champion EW. Functional recovery after hip fracture. *Arch Phys Med Rehabil* 1987 Oct;68(10):735-740. [Medline: [3662784](https://pubmed.ncbi.nlm.nih.gov/3662784/)]

36. Leibson CL, Tosteson ANA, Gabriel SE, Ransom JE, Melton LJ. Mortality, disability, and nursing home use for persons with and without hip fracture: a population-based study. *J Am Geriatr Soc* 2002 Oct 17;50(10):1644-1650. [doi: [10.1046/j.1532-5415.2002.50455.x](https://doi.org/10.1046/j.1532-5415.2002.50455.x)] [Medline: [12366617](https://pubmed.ncbi.nlm.nih.gov/12366617/)]
37. Cooper C, Atkinson EJ, Jacobsen SJ, O'Fallon WM, Melton LJ. Population-based study of survival after osteoporotic fractures. *Am J Epidemiol* 1993 May 01;137(9):1001-1005. [doi: [10.1093/oxfordjournals.aje.a116756](https://doi.org/10.1093/oxfordjournals.aje.a116756)] [Medline: [8317445](https://pubmed.ncbi.nlm.nih.gov/8317445/)]
38. Morin S, Lix LM, Azimae M, Metge C, Caetano P, Leslie WD. Mortality rates after incident non-traumatic fractures in older men and women. *Osteoporos Int* 2011 Sep 16;22(9):2439-2448. [doi: [10.1007/s00198-010-1480-2](https://doi.org/10.1007/s00198-010-1480-2)] [Medline: [21161507](https://pubmed.ncbi.nlm.nih.gov/21161507/)]
39. van Geel TACM, Eisman JA, Geusens PP, van den Bergh JPW, Center JR, Dinant GJ. The utility of absolute risk prediction using FRAX® and Garvan Fracture Risk Calculator in daily practice. *Maturitas* 2014 Feb;77(2):174-179. [doi: [10.1016/j.maturitas.2013.10.021](https://doi.org/10.1016/j.maturitas.2013.10.021)] [Medline: [24287178](https://pubmed.ncbi.nlm.nih.gov/24287178/)]
40. van Geel TA, van den Bergh JPW, Dinant GJ, Geusens PP. Individualizing fracture risk prediction. *Maturitas* 2010 Feb;65(2):143-148. [doi: [10.1016/j.maturitas.2009.12.007](https://doi.org/10.1016/j.maturitas.2009.12.007)] [Medline: [20053513](https://pubmed.ncbi.nlm.nih.gov/20053513/)]
41. Leslie WD, Lix LM. Comparison between various fracture risk assessment tools. *Osteoporos Int* 2014 Jan 25;25(1):1-21. [doi: [10.1007/s00198-013-2409-3](https://doi.org/10.1007/s00198-013-2409-3)] [Medline: [23797847](https://pubmed.ncbi.nlm.nih.gov/23797847/)]
42. Leslie WD, Berger C, Langsetmo L, Lix LM, Adachi JD, Hanley DA, Canadian Multicentre Osteoporosis Study Research Group. Construction and validation of a simplified fracture risk assessment tool for Canadian women and men: results from the CaMos and Manitoba cohorts. *Osteoporos Int* 2011 Jun 22;22(6):1873-1883 [FREE Full text] [doi: [10.1007/s00198-010-1445-5](https://doi.org/10.1007/s00198-010-1445-5)] [Medline: [20967422](https://pubmed.ncbi.nlm.nih.gov/20967422/)]
43. Hippisley-Cox J, Coupland C. Derivation and validation of updated QFracture algorithm to predict risk of osteoporotic fracture in primary care in the United Kingdom: prospective open cohort study. *BMJ* 2012 May 22;344(may22 1):e3427-e3427. [doi: [10.1136/bmj.e3427](https://doi.org/10.1136/bmj.e3427)] [Medline: [22619194](https://pubmed.ncbi.nlm.nih.gov/22619194/)]
44. Kanis JA, Johnell O, Oden A, Johansson H, McCloskey E. FRAX and the assessment of fracture probability in men and women from the UK. *Osteoporos Int* 2008 Apr 22;19(4):385-397 [FREE Full text] [doi: [10.1007/s00198-007-0543-5](https://doi.org/10.1007/s00198-007-0543-5)] [Medline: [18292978](https://pubmed.ncbi.nlm.nih.gov/18292978/)]
45. NOGG 2017: Clinical guideline for the prevention and treatment of osteoporosis. National Osteoporosis Guideline Group. 2017. URL: <https://www.sheffield.ac.uk/NOGG/NOGG%20Guideline%202017.pdf> [accessed 2020-01-05]
46. Lewiecki EM, Laster AJ, Miller PD, Bilezikian JP. More bone density testing is needed, not less. *J Bone Miner Res* 2012 Apr 20;27(4):739-742 [FREE Full text] [doi: [10.1002/jbmr.1580](https://doi.org/10.1002/jbmr.1580)] [Medline: [22383457](https://pubmed.ncbi.nlm.nih.gov/22383457/)]
47. Siris ES, Boonen S, Mitchell PJ, Bilezikian J, Silverman S. What's in a name? What constitutes the clinical diagnosis of osteoporosis? *Osteoporos Int* 2012 Aug 28;23(8):2093-2097. [doi: [10.1007/s00198-012-1991-0](https://doi.org/10.1007/s00198-012-1991-0)] [Medline: [22543575](https://pubmed.ncbi.nlm.nih.gov/22543575/)]
48. Chen Y, Miller PD, Barrett-Connor E, Weiss TW, Sajjan SG, Siris ES. An approach for identifying postmenopausal women age 50-64 years at increased short-term risk for osteoporotic fracture. *Osteoporos Int* 2007 Sep 27;18(9):1287-1296. [doi: [10.1007/s00198-007-0380-6](https://doi.org/10.1007/s00198-007-0380-6)] [Medline: [17464525](https://pubmed.ncbi.nlm.nih.gov/17464525/)]
49. Miller PD, Barlas S, Brenneman SK, Abbott TA, Chen Y, Barrett-Connor E, et al. An approach to identifying osteopenic women at increased short-term risk of fracture. *Arch Intern Med* 2004 May 24;164(10):1113-1120. [doi: [10.1001/archinte.164.10.1113](https://doi.org/10.1001/archinte.164.10.1113)] [Medline: [15159269](https://pubmed.ncbi.nlm.nih.gov/15159269/)]
50. Black DM, Steinbuch M, Palermo L, Dargent-Molina P, Lindsay R, Hoseyni MS, et al. An assessment tool for predicting fracture risk in postmenopausal women. *Osteoporos Int* 2001 Aug 1;12(7):519-528. [doi: [10.1007/s001980170072](https://doi.org/10.1007/s001980170072)] [Medline: [11527048](https://pubmed.ncbi.nlm.nih.gov/11527048/)]
51. Ferizi U, Honig S, Chang G. Artificial intelligence, osteoporosis and fragility fractures. *Current Opinion in Rheumatology* 2019;31(4):368-375. [doi: [10.1097/bor.0000000000000607](https://doi.org/10.1097/bor.0000000000000607)]
52. Kruse C, Eiken P, Vestergaard P. Machine Learning Principles Can Improve Hip Fracture Prediction. *Calcif Tissue Int* 2017 Apr 14;100(4):348-360. [doi: [10.1007/s00223-017-0238-7](https://doi.org/10.1007/s00223-017-0238-7)] [Medline: [28197643](https://pubmed.ncbi.nlm.nih.gov/28197643/)]
53. Kim S, Yoo T, Oh E, Kim D. Osteoporosis risk prediction using machine learning and conventional methods. *Conf Proc IEEE Eng Med Biol Soc. - 2013*;188:191. [doi: [10.1109/embc.2013.6609469](https://doi.org/10.1109/embc.2013.6609469)] [Medline: [24109656](https://pubmed.ncbi.nlm.nih.gov/24109656/)]
54. Wright NC, Daigle SG, Melton ME, Delzell ES, Balasubramanian A, Curtis JR. The design and validation of a new algorithm to identify incident fractures in administrative claims data. *J Bone Miner Res* 2019 Oct;34(10):1798-1807 [FREE Full text] [doi: [10.1002/jbmr.3807](https://doi.org/10.1002/jbmr.3807)] [Medline: [31170317](https://pubmed.ncbi.nlm.nih.gov/31170317/)]
55. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. *arXiv.org* 2013 Sep 07:1-12 [FREE Full text]
56. McInnes L, Healy J, Saul N, Großberger L. UMAP: Uniform Manifold Approximation and Projection. *JOSS* 2018 Sep;3(29):861. [doi: [10.21105/joss.00861](https://doi.org/10.21105/joss.00861)]
57. Nowak J, Taspinar A, Scherer R. LSTM Recurrent Neural Networks for Short Text and Sentiment Classification. In: Rutkowski L, Korytkowski M, Scherer R, Tadeusiewicz R, Zadeh L, Zurada J, editors. *Artificial Intelligence and Soft Computing. 16th International Conference, ICAISC 2017, Zakopane, Poland, June 11-15, 2017, Proceedings, Part II*. Cham: Springer International Publishing; Jun 15, 2017:553-562.
58. Le QV, Mikolov T. Distributed representations of sentences and documents. *Stanford University Quoc Le Profile*. URL: https://cs.stanford.edu/~quocle/paragraph_vector.pdf [accessed 2020-10-07]

59. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.: ACM Press; 2016 Aug Presented at: International Conference on Knowledge Discovery and Data Mining (KDD); August 13-17, 2020; San Francisco p. 785-794 URL: <https://dl.acm.org/doi/10.1145/2939672.2939785> [doi: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785)]
60. Cosman F, de Beur SJ, LeBoff MS, Lewiecki EM, Tanner B, Randall S, National Osteoporosis Foundation. Clinician's guide to prevention and treatment of osteoporosis. *Osteoporos Int* 2014 Oct 15;25(10):2359-2381 [FREE Full text] [doi: [10.1007/s00198-014-2794-2](https://doi.org/10.1007/s00198-014-2794-2)] [Medline: [25182228](https://pubmed.ncbi.nlm.nih.gov/25182228/)]
61. Eastell R, Rosen CJ, Black DM, Cheung AM, Murad MH, Shoback D. Pharmacological Management of Osteoporosis in Postmenopausal Women: An Endocrine Society* Clinical Practice Guideline. *J Clin Endocrinol Metab* 2019 May 01;104(5):1595-1622. [doi: [10.1210/jc.2019-00221](https://doi.org/10.1210/jc.2019-00221)] [Medline: [30907953](https://pubmed.ncbi.nlm.nih.gov/30907953/)]
62. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ* 2020 Mar 25:m689. [doi: [10.1136/bmj.m689](https://doi.org/10.1136/bmj.m689)]
63. Hodsman AB, Bauer DC, Dempster DW, Dian L, Hanley DA, Harris ST, et al. Parathyroid hormone and teriparatide for the treatment of osteoporosis: a review of the evidence and suggested guidelines for its use. *Endocr Rev* 2005 Aug;26(5):688-703. [doi: [10.1210/er.2004-0006](https://doi.org/10.1210/er.2004-0006)] [Medline: [15769903](https://pubmed.ncbi.nlm.nih.gov/15769903/)]
64. Forteo Package Insert. Lilly USA, LLC. 2020 Apr 06. URL: <https://pi.lilly.com/us/forteo-pi.pdf> [accessed 2020-05-11]
65. Neer RM, Arnaud CD, Zanchetta JR, Prince R, Gaich GA, Reginster JY, et al. Effect of parathyroid hormone (1-34) on fractures and bone mineral density in postmenopausal women with osteoporosis. *N Engl J Med* 2001 May 10;344(19):1434-1441. [doi: [10.1056/NEJM200105103441904](https://doi.org/10.1056/NEJM200105103441904)] [Medline: [11346808](https://pubmed.ncbi.nlm.nih.gov/11346808/)]
66. Lindsay R, Krege JH, Marin F, Jin L, Stepan JJ. Teriparatide for osteoporosis: importance of the full course. *Osteoporos Int* 2016 Aug;27(8):2395-2410 [FREE Full text] [doi: [10.1007/s00198-016-3534-6](https://doi.org/10.1007/s00198-016-3534-6)] [Medline: [26902094](https://pubmed.ncbi.nlm.nih.gov/26902094/)]
67. Tymlos Package Insert. Radius Health, Inc. 2018 Oct. URL: <https://radiuspharm.com/wp-content/uploads/tymlos/tymlos-prescribing-information.pdf> [accessed 2020-05-10]
68. Miller PD, Hattersley G, Riis BJ, Williams GC, Lau E, Russo LA, ACTIVE Study Investigators. Effect of abaloparatide vs placebo on new vertebral fractures in postmenopausal women with osteoporosis: a randomized clinical trial. *JAMA* 2016 Aug 16;316(7):722-733. [doi: [10.1001/jama.2016.11136](https://doi.org/10.1001/jama.2016.11136)] [Medline: [27533157](https://pubmed.ncbi.nlm.nih.gov/27533157/)]
69. Kendler DL, Marin F, Zerbini CAF, Russo LA, Greenspan SL, Zikan V, et al. Effects of teriparatide and risedronate on new fractures in post-menopausal women with severe osteoporosis (VERO): a multicentre, double-blind, double-dummy, randomised controlled trial. *Lancet* 2018 Jan 20;391(10117):230-240. [doi: [10.1016/S0140-6736\(17\)32137-2](https://doi.org/10.1016/S0140-6736(17)32137-2)] [Medline: [29129436](https://pubmed.ncbi.nlm.nih.gov/29129436/)]
70. Ballane G, Cauley JA, Luckey MM, El-Hajj Fuleihan G. Worldwide prevalence and incidence of osteoporotic vertebral fractures. *Osteoporos Int* 2017 Feb 6;28(5):1531-1542. [doi: [10.1007/s00198-017-3909-3](https://doi.org/10.1007/s00198-017-3909-3)]

Abbreviations

AUPRC: area under the precision-recall curve

AUROC: area under the receiver operating characteristics curve

FRAX: University of Sheffield Fracture Risk Assessment Tool

GIH-BFRC: Garvan Institute of Health Bone Fracture Risk Calculator

ICD: International Classification of Diseases

Edited by G Eysenbach; submitted 15.07.20; peer-reviewed by C Fincham, M Pradhan; comments to author 06.08.20; revised version received 05.09.20; accepted 12.09.20; published 16.10.20

Please cite as:

Almog YA, Rai A, Zhang P, Moulaison A, Powell R, Mishra A, Weinberg K, Hamilton C, Oates M, McCloskey E, Cummings SR
Deep Learning With Electronic Health Records for Short-Term Fracture Risk Identification: Crystal Bone Algorithm Development and Validation

J Med Internet Res 2020;22(10):e22550

URL: <http://www.jmir.org/2020/10/e22550/>

doi: [10.2196/22550](https://doi.org/10.2196/22550)

PMID:

©Yasmeen Adar Almog, Angshu Rai, Patrick Zhang, Amanda Moulaison, Ross Powell, Anirban Mishra, Kerry Weinberg, Celeste Hamilton, Mary Oates, Eugene McCloskey, Steven R Cummings. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 16.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution

License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.