Development and validation of sex-specific hip fracture prediction models using electronic health records: a retrospective, population-based cohort study

Gloria Hoi-Yee Li,[a] Ching-Lung Cheung,[b,*] Kathryn Choon-Beng Tan,[c] Annie Wai-Chee Kung,[c] Timothy Chi-Yui Kwok,[d] Wallis Cheuk-Yin Lau,[e] Janus Siu-Him Wong,[f] Warrington W. Q. Hsu,[b] Christian Fang,[f] and Ian Chi-Kei Wong[b,e,g,h]

[a]Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong SAR, China
[b]Department of Pharmacology and Pharmacy, The University of Hong Kong, Hong Kong SAR, China
[c]Department of Medicine, School of Clinical Medicine, The University of Hong Kong, Hong Kong SAR, China
[d]Department of Medicine & Therapeutics and School of Public Health, The Chinese University of Hong Kong, Hong Kong SAR, China
[e]School of Pharmacy, University College London, London, UK
[f]Department of Orthopaedics and Traumatology, School of Clinical Medicine, The University of Hong Kong, Hong Kong SAR, China
[g]Musketeers Foundation Institute of Data Science, The University of Hong, Hong Kong SAR, China
[h]Aston School of Pharmacy, Aston University, Birmingham B4 7ET, UK

Summary

Background Hip fracture is associated with immobility, morbidity, mortality, and high medical cost. Due to limited availability of dual-energy X-ray absorptiometry (DXA), hip fracture prediction models without using bone mineral density (BMD) data are essential. We aimed to develop and validate 10-year sex-specific hip fracture prediction models using electronic health records (EHR) without BMD.

Methods In this retrospective, population-based cohort study, anonymized medical records were retrieved from the Clinical Data Analysis and Reporting System for public healthcare service users in Hong Kong aged ≥60 years as of 31 December 2005. A total of 161,051 individuals (91,926 female; 69,125 male) with complete follow-up from 1 January 2006 till the study end date on 31 December 2015 were included in the derivation cohort. The sex-stratified derivation cohort was randomly divided into 80% training and 20% internal testing datasets. An independent validation cohort comprised 3046 community-dwelling participants aged ≥60 years as of 31 December 2005 from the Hong Kong Osteoporosis Study, a prospective cohort which recruited participants between 1995 and 2010. With 395 potential predictors (age, diagnosis, and drug prescription records from EHR), 10-year sex-specific hip fracture prediction models were developed using stepwise selection by logistic regression (LR) and four machine learning (ML) algorithms (gradient boosting machine, random forest, eXtreme gradient boosting, and single-layer neural networks) in the training cohort. Model performance was evaluated in both internal and independent validation cohorts.

Findings In female, the LR model had the highest AUC (0.815; 95% Confidence Interval [CI]: 0.805–0.825) and adequate calibration in internal validation. Reclassification metrics showed the LR model had better discrimination and classification performance than the ML algorithms. Similar performance was attained by the LR model in independent validation, with high AUC (0.841; 95% CI: 0.807–0.87) comparable to other ML algorithms. In internal validation for male, LR model had high AUC (0.818; 95% CI: 0.801–0.834) and it outperformed all ML models as indicated by reclassification metrics, with adequate calibration. In independent validation, the LR model had high AUC (0.898; 95% CI: 0.857–0.939) comparable to ML algorithms. Reclassification metrics demonstrated that LR model had the best discrimination performance.

Interpretation Even without using BMD data, the 10-year hip fracture prediction models developed by conventional LR had better discrimination performance than the models developed by ML algorithms. Upon further validation in independent cohorts, the LR models could be integrated into the routine clinical workflow, aiding the identification of people at high risk for DXA scan.

*Corresponding author. Department of Pharmacology and Pharmacy, The University of Hong Kong, Pokfulam, Hong Kong.
    E-mail address: lung1212@hku.hk (C.-L. Cheung).

Keywords: Hip fracture; Prediction model; Machine learning

Research in context

Evidence before this study

We searched PubMed for hip fracture prediction tools developed from 2012 to 2022 using the search terms ("hip fracture") AND (predict* OR assess*) AND (tool OR model). The majority of the identified studies evaluated the importance of some risk factors in hip fracture development in community-dwelling cohorts or patient subgroups. Some studies evaluated the accuracy of existing models (mainly FRAX) in patients with a particular disease. Three studies adopted machine learning approach to predict future fracture risk, but only conventional bone-related risk factors were used in developing the model. BMD data was used as a predictor in two of these studies.

Added value of this study

In this population-based study using stepwise selection by logistic regression and four machine learning (ML) algorithms, 10-year sex-specific hip fracture prediction models were developed using age, all diagnosis and drug prescription records as predictors, which were retrieved from a representative electronic medical database in Hong Kong. The discrimination and calibration performance were evaluated in both the internal testing cohort and independent validation cohort which comprised community-dwelling individuals. Without using BMD data and other clinical parameters such as height and weight, the logistic regression model had high discrimination performance and outperformed all ML models in both female and male. Adequate calibration was also observed for female.

Implications of all the available evidence

Using electronic medical records as the only predictors, logistic regression models performed better than ML algorithms in predicting the 10-year hip fracture risk in both female and male. Upon further validation, the logistic regression models may be integrated to the routine clinical workflow. These prediction models may be applied at both public healthcare service setting and the community-dwelling individuals at population-level, aiding to triage individuals who are at high risk of hip fracture for prioritized DXA scan, and subsequent treatment initiation. Such measures are expected to facilitate early prevention, timely diagnosis, and treatment of osteoporosis.

Introduction

Osteoporosis is a prevalent disease characterized by low bone mass and deterioration in bone strength and microarchitecture, which leads to increased risk of fragility fracture. Among all fragility fractures, hip fracture is known to be associated with high immobility, morbidity, and mortality. Earlier projection in 1990s demonstrated that there will be around 4.5–6.26 million hip fractures globally by 2050, with half of them from Asia.[1,2] This concurs with our recent projection that the number of hip fracture in Asia will reach 2.56 million in 2050, leading to an annual direct medical cost of around USD15 billion in Asia.[3] Given that hip fracture is associated with high medical cost, prevention of hip fracture is not only essential for people at high risk and their caregivers, but also the healthcare system and society.

Dual-energy X-ray absorptiometry (DXA) is the gold standard for measurement of bone mineral density (BMD) and diagnosis of osteoporosis. It is also an important facility to predict fracture. Yet, its availability is considerably low, especially in the developing countries and regions.[4] Even a majority of European countries had insufficient provision of DXA machines for the general population to meet the requirements of practice guidelines.[5] The average waiting time for DXA scan in European countries could be as long as 180 days.[5] Due to the limited resources for DXA scan services, it is important to develop a fracture risk prediction model without BMD data as a routine screening tool in public healthcare setting, which facilitates the prioritization of people at high risk for DXA scan, aiding early diagnosis and timely treatment of osteoporosis.

Existing prediction tools, such as FRAX, were developed using data mainly from Caucasians.[6] We previously found that ethnic-specific clinical risk factors outperformed the performance of FRAX in Hong Kong,[7] demonstrating the importance of developing a population-specific hip fracture prediction tool. Recently, machine learning (ML) algorithms were applied to develop fracture risk prediction models.[8–10] Notably, most ML models were developed among people in Europe and United States, mainly used to predict the short-term fracture risk in up to 5 years.[8–10] In this study, we aimed to develop and validate models that predict the 10-year risk of hip fracture for individuals in Hong Kong using age, diagnosis and drug prescription data in the form of electronic health records (EHR), but in the absence of conventional clinical parameters such as BMD, height, weight and body mass index (BMI). To account for sex-specific factors contributing to the different causes of osteoporosis and hip fracture incidence between the two sexes, these

prediction models were separately developed and validated in female and male.

Methods

Study design and participants

In this retrospective, population-based cohort study, anonymized medical records were retrieved from the Clinical Data Analysis and Reporting System (CDARS), a large and representative electronic medical database in Hong Kong managed by the Hong Kong Hospital Authority (HA). The HA is a public healthcare service provider that manages 43 hospitals and institutions, and 122 outpatient clinics, serving >80% of hospital admissions. Approximately 98% of hip fracture in Hong Kong was admitted to HA hospitals,[11] and the hip fracture coding in CDARS was previously validated with a positive predictive value (PPV) of 100%,[12] suggesting that CDARS data is representative and accurate, particularly for hip fracture. The medical records available in CDARS comprise demographics, prescription (British National Formulary [BNF]), diagnosis (International Classification of Disease, 9th revision, Clinical Modification [ICD-9-CM]), admission, procedures, and laboratory tests.

Fig. 1 illustrates the study design. As of 31 December 2005 (index date), about 740,000 public healthcare service users aged ≥60 had admission records at in-patient, out-patient, or accident & emergency services from 1 January to 31 December 2005 in CDARS. Approximately one-third of them were randomly selected, and they were representative of the targeted population based on the demographics (Supplementary Fig. S1). Individuals with complete follow-up from 1 January 2006 till the study end date on 31 December 2015 were included in the derivation cohort. The outcome of interest was the 10-year risk of developing hip fracture, which was identified by ICD-9-CM code of 820.xx.[12] The derivation cohort was sex-stratified, and each sex-specific sub-cohort was randomly split into the training (80%) and internal testing (20%) datasets. Conventional statistical model and ML algorithms were used to develop the prediction models in the training dataset, followed by validation in the internal testing dataset. Performance of the prediction models were further assessed in the independent validation cohort comprising participants aged ≥60 from the Hong Kong Osteoporosis Study (HKOS), which was described elsewhere.[13] Briefly, the HKOS comprised >9000 community-dwelling Southern Chinese participants, who were followed using EHR from CDARS. The independent validation cohort comprised 3048 HKOS participants aged ≥60 as of 31 December 2005, without overlap with the derivation cohort. The study adhered to the reporting guidelines of developing and validating a prediction model as stated in the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD).

The study protocol was approved by the institutional review board of the University of Hong Kong and the HA Hong Kong West Cluster (reference: UW 19-798), and the Hong Kong Polytechnic University (reference: HSEARS20201109004). As the EHR from CDARS were anonymized, relevant regulations in Hong Kong did not require the informed consent from study participants. For the independent validation cohort, all the participants gave informed consent to participate in the HKOS at their baseline visit.

Predictor variables

Potential predictors, including age on index date, all diagnosis and drug prescription records within one year preceding the index date, were retrieved from CDARS for individuals in the derivation and independent validation cohorts. The presence or absence of each diagnosis code (as sub-chapters of ICD-9-CM) was recorded as binary coding using the icd package[14] in R. Whether an individual was prescribed a class of drug (as BNF codes including chapters and sections) was also recorded as binary coding. Out of 395 potential predictors, 163 diagnosis and drug prescription variables with zero or near-zero variance (binary variables with ≤0.1% prevalence in the sex-stratified cohort) were excluded, leaving 232 potential predictor variables for the female and male derivation cohorts respectively (Supplementary Table S1) to train the prediction models. Age was the only continuous predictor variable. One-sample Kolmogorov–Smirnov test showed that age did not follow a normal distribution and it was presented as median (interquartile range). Age between groups were compared using Kruskal–Wallis test. For other predictor variables which are all binary, data are presented as numbers (percentage), and comparison between groups was done using chi-square test.

Development of prediction models

For the conventional statistical model, all potential predictors were included at the start, followed by a stepwise selection by logistic regression (LR) which added and dropped predictors to identify a model with the lowest Akaike Information Criteria (AIC),[15] penalizing addition of variables into the model. An R package, "MASS", was employed to implement the stepwise algorithm for LR.[16] Four ML algorithms (including gradient boosting machine [GBM], random forest [RF], eXtreme gradient boosting [xgbTree], and neural networks with a single hidden layer [nnet]) were adopted to train the prediction model, utilizing the caret package in R.[17] For each algorithm, hyperparameters were optimized with 10 repeats of 10-fold cross-validation to maximize the area under the receiver operating characteristic (ROC) curve (AUC) of the training model. The final hyperparameters used in the prediction models were listed in Supplementary Table S2.
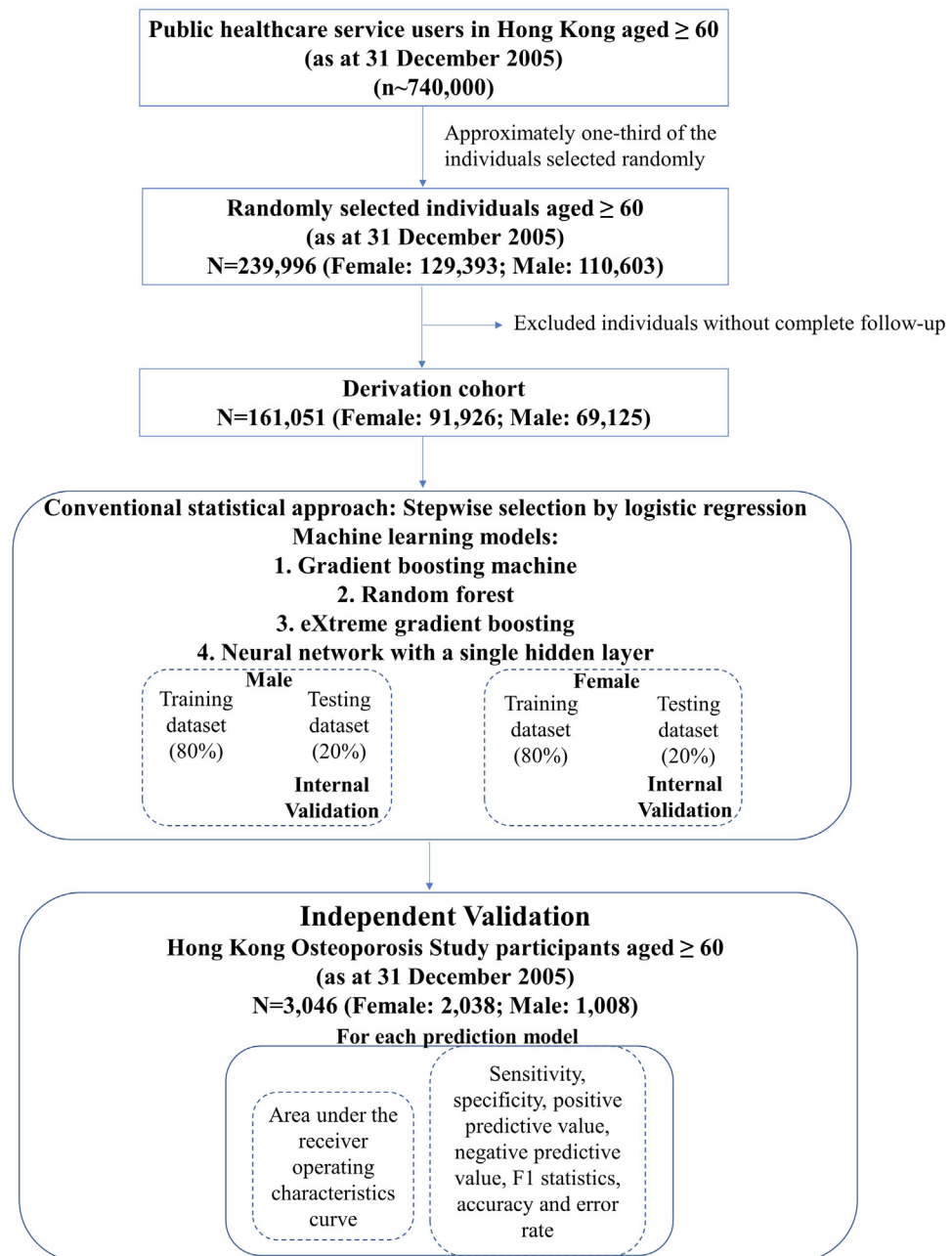
Public healthcare service users in Hong Kong aged ≥ 60
(as at 31 December 2005)
(n~740,000)

Approximately one-third of the individuals selected randomly

Randomly selected individuals aged ≥ 60
(as at 31 December 2005)
N=239,996 (Female: 129,393; Male: 110,603)

Excluded individuals without complete follow-up

Derivation cohort
N=161,051 (Female: 91,926; Male: 69,125)

Conventional statistical approach: Stepwise selection by logistic regression
Machine learning models:
1. Gradient boosting machine
2. Random forest
3. eXtreme gradient boosting
4. Neural network with a single hidden layer

Male
Training dataset (80%) | Testing dataset (20%)
Internal Validation

Female
Training dataset (80%) | Testing dataset (20%)
Internal Validation

Independent Validation
Hong Kong Osteoporosis Study participants aged ≥ 60
(as at 31 December 2005)
N=3,046 (Female: 2,038; Male: 1,008)
For each prediction model

Area under the receiver operating characteristics curve

Sensitivity, specificity, positive predictive value, negative predictive value, F1 statistics, accuracy and error rate

Fig. 1: Study design and workflow of cohort derivation.

Evaluation of prediction models

The general diagnostic accuracy of each model was evaluated by the AUC in the internal testing and independent validation datasets. The optimal cut-off value for hip fracture risk classification was determined based on the ROC analysis of the training dataset using the Youden's index.[18] The sensitivity, specificity, PPV, negative predictive value (NPV), F1 statistics, accuracy and error rate were evaluated for each prediction model in the internal testing and independent validation cohorts. DeLong's test was used to compare the AUC of two models. With the LR model as reference, whether the ML algorithms had improvement in discrimination performance were assessed using the category-less net reclassification index (NRI) and integrated discrimination improvement index (IDI), which were computed using the Hmisc package[19] in R. As a measure of both discrimination and calibration,[20] the Brier score was calculated as the mean squared

error between the actual event (fracture) and estimated probability.[21] The calibration slope, intercept, and the Spiegelhalter Z-test (with perfect calibration as the null hypothesis)[22] were computed using the rms package[23] in R. Smaller Brier score, insignificant Spiegelhalter Z-test, a calibration slope closer to 1 and intercept closer to 0 imply better calibration. The observed and predicted probability of different models in internal and independent validation were presented as calibration curves.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, and data interpretation, or writing of the report. All authors had full access to all the data in the study and accepted the responsibility to submit for publication.

Results

Cohort participants

Fig. 1 outlines the workflow in selecting individuals included in the derivation cohorts. The derivation cohort comprised 161,051 individuals (91,926 female; 69,125 male). Their baseline characteristics are presented in Tables 1 and 2. The proportion of hip fracture cases in the derivation cohort was preserved in the constituting training and testing cohorts. In the female derivation cohort, 10.3% of the individuals had hip fracture within the 10-year follow-up (Table 1). Only 6% individuals in the male derivation cohort had hip fracture events within the follow-up period (Table 2). The baseline characteristics within one year prior to index date were similar among individuals in the training and internal testing cohorts (Tables 1 and 2). The independent validation cohort from the HKOS comprised a total of 3046 community-dwelling individuals (2038 female; 1008 male), with more female (66.91%) than the derivation cohort (57.08%). Individuals in the independent validation cohort were younger (for female) and had fewer hip fracture cases during follow-up (Tables 1 and 2). Some known risk factors of fracture, such as diagnosis/drug prescription records of cardiovascular disease (CVD), diabetes, rheumatic diseases and gout, were less prevalent in the independent validation cohort.

Characteristics	Derivation cohort		Independent validation cohort n = 2038
	Training cohort n = 73,541	Testing cohort n = 18,385[a]	
Hip fracture cases within 10-year follow-up[b]	7568 (10.29)	1892 (10.29)	145 (7.11)[e]
Age on index date[c]	71 [65.88–76.96]	71 [65.74–76.6]	68.99 [64.97–74.86][e]
Medical history (within 1-year prior to index date)			
Disease of the cardiovascular system			
Diagnosis record[d]	3858 (5.25)	987 (5.37)	87 (4.27)[e]
Drug prescription record (BNF: 2.x)	43,312 (58.9)	10,806 (58.78)	953 (46.76)[e]
With diagnosis and/or drug prescription records	**43,513 (59.17)**	**10,862 (59.08)**	**971 (47.64)[e]**
Chronic obstructive pulmonary disease and allied conditions			
Diagnosis record (ICD-9-CM: 490.xx-496.xx)	574 (0.78)	130 (0.71)	10 (0.49)
Drug prescription record (BNF: 3.1 and 3.2)	4617 (6.28)	1089 (5.92)	68 (3.34)[e]
With diagnosis and/or drug prescription records	**4712 (6.41)**	**1108 (6.03)**	**72 (3.53)[e]**
Diabetes			
Diagnosis record (ICD-9-CM: 250.xx)	1186 (1.61)	328 (1.78)	22 (1.08)[e]
Drug prescription record (BNF: 6.1)	13,026 (17.7)	3264 (17.86)	181 (8.88)[e]
With diagnosis and/or drug prescription records	**13,158 (17.89)**	**3323 (18.07)**	**186 (9.13)[e]**
Rheumatic diseases and gout			
Diagnosis record (ICD-9-CM: 274.xx, 725.xx-729.xx)	555 (0.75)	160 (0.87)	13 (0.64)
Drug prescription record (BNF: 10.1)	13,516 (18.38)	3380 (18.38)	270 (13.25)[e]
With diagnosis and/or drug prescription records	**13,679 (18.6)**	**3424 (18.62)**	**278 (13.64)[e]**
Dementia (including Alzheimer's disease)			
Diagnosis record (ICD-9-CM: 290.xx and 331.0)	227 (0.31)	59 (0.32)	2 (0.1)
Drug prescription record (BNF 4.11)	210 (0.29)	45 (0.24)	1 (0.05)
With diagnosis and/or drug prescription records	**385 (0.52)**	**93 (0.51)**	**3 (0.15)[e]**

BNF: British National Formulary; ICD-9-CM: International Classification of Disease, 9th revision, Clinical Modification. The bold values indicate the number (%) of individuals with diagnosis and/or drug prescription of the diseases, in contrast to the previous rows that indicate the number of either diagnosis or drug prescription of the diseases. [a]No significant difference observed between training and testing cohort (All p > 0.05). [b]The overall proportion of hip fracture cases were preserved in the 80% training and 20% testing cohorts. [c]Age on index date is presented as median [interquartile range]. All other binary variables are presented as numbers (percentage). [d]ICD-9-CM codes for cardiovascular disease: 390.xx-398.xx, 401.xx-405.xx, 410.xx-414.xx, 415.xx-417.xx, 420.xx-429.xx, 430.xx-438.xx, 440.xx-449.xx, 421.xx-459.xx. [e]Significant difference observed between derivation cohort and independent validation cohort (p < 0.05).

Table 1: Characteristics of the cohort participants in primary analysis. Characteristics of female cohort participants in the prediction model of 10-year risk of hip fracture.

Characteristics	Derivation cohort		Independent validation cohort n = 1008
	Training cohort n = 55,301	Testing cohort n = 13,824[a]	
Hip fracture cases within 10-year follow-up[b]	3301 (6.0)	825 (6.0)	36 (3.6)[e]
Age on index date[c]	69 [64.83–74.2]	69 [64.96–74.14]	68.84 [65.13–73.25]
Medical history (Within 1-year prior to index date)			
Disease of the cardiovascular system			
Diagnosis record[d]	3398 (6.14)	868 (6.28)	39 (3.87)[e]
Drug prescription record (BNF: 2.x)	31,074 (56.19)	7690 (55.63)	470 (46.63)[e]
With diagnosis and/or drug prescription records	**31,318 (56.63)**	**7751 (56.07)**	**474 (47.02)[e]**
Chronic obstructive pulmonary disease and allied conditions			
Diagnosis record (ICD-9-CM: 490.xx-496.xx)	716 (1.29)	174 (1.26)	8 (0.79)
Drug prescription record (BNF:3.1-3.2)	4709 (8.52)	1154 (8.35)	31 (3.08)[e]
With diagnosis and/or drug prescription records	**4761 (8.61)**	**1166 (8.43)**	**32 (3.17)[e]**
Diabetes			
Diagnosis record (ICD-9-CM: 250.xx)	907 (1.64)	209 (1.51)	7 (0.69)[e]
Drug prescription record (BNF: 6.1)	8562 (15.48)	2158 (15.61)	76 (7.54)[e]
With diagnosis and/or drug prescription records	**8674 (15.68)**	**2175 (15.73)**	**81 (8.04)[e]**
Rheumatic diseases and gout			
Diagnosis record (ICD-9-CM: 274.xx, 725.xx-729.xx)	640 (1.16)	138 (1)	5 (0.5)
Drug prescription record (BNF 10.1)	9908 (17.92)	2511 (18.16)	124 (12.3)[e]
With diagnosis and/or drug prescription records	**10,005 (18.09)**	**2537 (18.35)**	**127 (12.6)[e]**
Dementia (including Alzheimer's disease)			
Diagnosis record (ICD-9-CM: 290.xx and 331.0)	61 (0.11)	18 (0.13)	0 (0)
Drug prescription record (BNF 4.11)	68 (0.12)	15 (0.11)	0 (0)
With diagnosis and/or drug prescription records	**120 (0.22)**	**32 (0.23)**	**0 (0)**

BNF: British National Formulary; ICD-9-CM: International Classification of Disease, 9th revision, Clinical Modification. The bold values indicate the number (%) of individuals with diagnosis and/or drug prescription of the diseases, in contrast to the previous rows that indicate the number of either diagnosis or drug prescription of the diseases. [a]No significant difference observed between training and testing cohort (All p > 0.05). [b]The overall proportion of hip fracture cases were preserved in the 80% training and 20% testing cohorts. [c]Age on index date is presented as median [interquartile range]. All other binary variables are presented as numbers (percentage). [d]ICD-9-CM codes for cardiovascular disease: 390.xx-398.xx, 401.xx-405.xx, 410.xx-414.xx, 415.xx-417.xx, 420.xx-429.xx, 430.xx-438.xx, 440.xx-449.xx, 421.xx-459.xx. [e]Significant difference observed between derivation cohort and independent validation cohort (p < 0.05).

Table 2: Characteristics of the cohort participants in primary analysis. Characteristics of male cohort participants in the prediction model of 10-year risk of hip fracture.

Performance of prediction models

The discrimination performance metrics of the female prediction models in internal and independent validation cohorts are presented in Table 3. In the internal validation cohort, the stepwise selection by LR, GBM and xgbTree models attained the highest AUC of 0.815 (95% Confidence Interval: 0.805–0.825). Using the Youden's index to determine the optimal threshold for hip fracture classification, the LR model had moderate sensitivity and specificity (>0.7) (Table 3 and Supplementary Fig. S2). All the ML algorithms had statistically significant and negative IDI and NRI with reference to the LR model. The negative IDI indicated that the ML algorithms had lower integrated sensitivity and integrated specificity than the LR model, while the negative NRI implied that the reclassification of hip fracture and non-hip fracture cases made by the ML algorithms were worse than the LR model (Supplementary Table S3). The DeLong' test showed that the AUC of the LR model was significantly higher than the RF and nnet models (Table 3). The

LR model was well-calibrated, as suggested by the small Brier's score and insignificant Spiegelhalter Z-test (Supplementary Table S4 and Supplementary Fig. S3). In independent validation, the LR model attained a high AUC of 0.841 (0.807–0.87). With the threshold defined by the Youden's index, the LR model also had moderate sensitivity (0.69) but high specificity (0.82). Its AUC was significantly higher than the RF model, but comparable to other ML models with AUC in the range of 0.832–0.845 (Table 3 and Supplementary Fig. S4). The statistically significant and negative IDI and NRI showed that the LR model had better discrimination performance than the ML models (Supplementary Table S3). Due to the similar AUC of LR with GBM and xgbTree models, sensitivity analysis was performed to evaluate the reclassification metrics using these two gradient boosting models as reference (Supplementary Table S5). The LR model also had adequate calibration in independent validation (Supplementary Table S4 and Supplementary Fig. S5). The RF model was overfitted to the training cohort, and it

Algorithm used in model development	Stepwise selection by logistic regression	Gradient boosting machine	Random forest	eXtreme gradient boosting	Neural networks with a single hidden layer
Derivation cohort					
Training cohort					
AUC (95% CI)	0.823 (0.818–0.827)	0.823 (0.818–0.828)	0.996 (0.996–0.997)	0.826 (0.821–0.831)	0.825 (0.82–0.83)
Testing cohort					
AUC (95% CI)	0.815 (0.805–0.825)	0.815 (0.805–0.825)	0.78 (0.769–0.791)	0.815 (0.805–0.825)	0.803 (0.792–0.813)
Sensitivity	0.721	0.754	0.5	0.757	0.724
Specificity	0.754	0.724	0.868	0.721	0.739
PPV	0.252	0.239	0.302	0.237	0.241
NPV	0.959	0.962	0.938	0.963	0.959
F1	0.373	0.362	0.376	0.361	0.362
Accuracy	0.751	0.727	0.83	0.724	0.737
Error	0.249	0.273	0.17	0.276	0.263
Delong's test p-value	Reference	0.95	<0.0001	0.98	<0.0001
Independent validation cohort					
AUC (95% CI)	0.841 (0.807–0.87)	0.845 (0.811–0.879)	0.813 (0.779–0.848)	0.842 (0.808–0.877)	0.832 (0.797–0.867)
Sensitivity	0.69	0.724	0.51	0.731	0.731
Specificity	0.817	0.802	0.895	0.797	0.793
PPV	0.224	0.219	0.271	0.216	0.213
NPV	0.972	0.974	0.96	0.975	0.975
F1	0.338	0.336	0.354	0.333	0.33
Accuracy	0.808	0.796	0.868	0.792	0.789
Error	0.192	0.204	0.133	0.208	0.211
Delong's test p-value	Reference	0.15	0.016	0.69	0.29

AUC: area under the receiver operating characteristic curve; CI: confidence interval; NPV: negative predictive value; PPV: positive predictive value.

Table 3: **Discrimination performance of hip fracture risk prediction models for female.**

cannot be generalized to both the internal and independent validation cohorts (Table 3), as suggested by its poor calibration (Supplementary Table S4, Supplementary Figs. S3 and S5).

The discrimination performance of the prediction models developed for male are presented in Table 4. In internal validation, although the xgbTree model had a significantly higher AUC of 0.825 (0.809–0.84) than the LR model (0.818 [0.801–0.834]) (Table 4, Supplementary Fig. S6), the discrimination performance of LR model outperformed other models as indicated by the negative IDI and NRI of the ML models (Supplementary Table S6). Adequate calibration was also observed for the LR model (Supplementary Table S7 and Supplementary Fig. S7). In independent validation, the LR model had a high AUC of 0.898 (0.857–0.939), which was significantly higher than the RF model, but comparable to other ML models with AUC in the range of 0.898–0.905 (Table 4, Supplementary Fig. S8). The IDI and NRI of the GBM, RF and xgbTree models were statistically significant and negative, implying that the LR model had better discrimination performance than these ML models (Supplementary Table S6). The negative IDI of the nnet model reached statistical significance, but not the NRI (Supplementary Table S6). The sensitivity analysis using the GBM and xgbTree models as reference

were provided in Supplementary Table S8. Moreover, the calibration was inadequate in independent validation for all the male prediction models (Supplementary Table S7 and Supplementary Fig. S9). The worst calibration was observed for the RF model in both internal and independent validation (Supplementary Table S7, Supplementary Figs. S7 and S9), suggesting that the model may be overfitted to the training cohort (Table 4).

Association of predictors with hip fracture
Since the discrimination performance of the LR model outperformed the ML models in both female and male in internal testing and independent validation, the 20 top predictors adopted by the LR model having the strongest association with hip fracture are listed in Tables 5 and 6. Eleven of them were among the top 20 in both the female and male prediction models.

Discussion
In the current study, we utilized EHR of >160,000 individuals from a population-based cohort to develop 10-year sex-specific hip fracture risk prediction models in Hong Kong, using both conventional statistical approach and ML algorithms. The prediction models were validated in the internal testing cohort of public

Algorithm used in model development	Stepwise selection by logistic regression	Gradient boosting machine	Random forest	eXtreme gradient boosting	Neural networks with a single hidden layer
Derivation cohort					
Training cohort					
AUC (95% CI)	0.826 (0.819–0.834)	0.825 (0.818–0.833)	0.996 (0.995–0.997)	0.834 (0.827–0.841)	0.826 (0.819–0.834)
Testing cohort					
AUC (95% CI)	0.818 (0.801–0.834)	0.824 (0.808–0.839)	0.775 (0.757–0.793)	0.825 (0.809–0.84)	0.818 (0.802–0.833)
Sensitivity	0.744	0.742	0.416	0.736	0.727
Specificity	0.749	0.753	0.923	0.758	0.761
PPV	0.158	0.16	0.254	0.162	0.162
NPV	0.979	0.979	0.961	0.978	0.978
F1	0.261	0.263	0.315	0.265	0.264
Accuracy	0.749	0.752	0.892	0.756	0.759
Error	0.251	0.248	0.108	0.244	0.241
Delong's test p-value	Reference	0.066	<0.0001	0.019	0.88
Independent validation cohort					
AUC (95% CI)	0.898 (0.857–0.939)	0.898 (0.857–0.939)	0.84 (0.783–0.896)	0.9 (0.861–0.939)	0.905 (0.863–0.947)
Sensitivity	0.806	0.75	0.25	0.75	0.806
Specificity	0.817	0.81	0.957	0.824	0.823
PPV	0.14	0.127	0.176	0.136	0.144
NPV	0.991	0.989	0.972	0.989	0.991
F1	0.239	0.218	0.207	0.231	0.245
Accuracy	0.817	0.808	0.932	0.821	0.822
Error	0.184	0.193	0.069	0.179	0.178
Delong's test p-value	Reference	0.99	0.0050	0.76	0.34

AUC: area under the receiver operating characteristic curve; CI: confidence interval; NPV: negative predictive value; PPV: positive predictive value.

Table 4: **Discrimination performance of hip fracture risk prediction models for male.**

healthcare service users, and the independent validation cohort of community-dwelling individuals. The conventional LR model outperformed the ML models in both female and male. In particular, the LR model for female was adequately calibrated, suggesting the potential usefulness clinically. To our knowledge, this is one of the largest samples used to develop hip fracture prediction models among the Asians.

One noticeable feature of our prediction models is that we included age, all diagnosis and drug prescription records from the electronic medical database as potential predictors, irrespective of their prior association with hip fracture. Most importantly, BMD data was not used in model development. Since the EHR was input by clinicians and healthcare professionals at patient visit, the readily available data enhanced the feasibility of integrating the prediction models into the routine clinical workflow of public healthcare setting in Hong Kong. Even in the absence of BMD data, the LR model for female had AUC >0.8 in both internal testing and independent validation. In addition to adequate calibration, this model is likely to be clinically useful in risk stratification.[24] Although the AUC of the LR model for male was also high in internal and independent validation (>0.8), the independent validation was inadequately calibrated, which may be attributed to the relatively small sample number of male participants in HKOS.

Further validation of the male prediction models in independent cohort of larger sample size is warranted to evaluate its potential usefulness in hip fracture risk prediction. In comparison with existing fracture prediction tools, such as QFracture,[25] FRAX[6] and Garvan,[26] they included only a pre-defined set of conventional risk factors of hip fracture in development of the prediction model. Notably, clinical parameters such as weight and/or height were used as the conventional predictors in FRAX[6] and Garvan,[26] if BMD data was unavailable. Conversely, our prediction models did not include any clinical parameters (such as weight, height, and BMD) as predictor. In addition, while the internal testing cohort consisted of public healthcare service users only, our independent validation cohort comprised the HKOS participants who were community-dwelling individuals, demonstrating the potentially high generalizability of our prediction models.

Several studies have adopted the ML approach to predict future fracture risk.[8–10] A study utilized the national Danish patient data of 6600 individuals to develop a 5-year hip fracture prediction model. With DXA data and laboratory tests, their prediction models had a good performance with AUC >0.9.[10] Nevertheless, DXA screening is not easily accessible,[4,5] limiting its generalizability. Another study used data of 5130 individuals from the Osteoporosis Fractures in Men (MrOS) for

Predictors	Training cohort		
	OR	(95% CI)	p-value
Age on index date	1.161	(1.156–1.165)	<0.0001
Diagnosis			
Accidental falls	1.673	(1.403–1.996)	<0.0001
Chronic obstructive pulmonary disease and allied Conditions	1.506	(1.181–1.921)	0.0010
Dorsopathies	1.415	(1.185–1.689)	0.0001
Nephritis, nephrotic syndrome, and nephrosis	2.299	(1.548–3.417)	<0.0001
Organic psychotic conditions	1.651	(1.278–2.133)	0.0001
Drug prescription			
Anaemias and some other blood disorders	1.52	(1.293–1.786)	<0.0001
Antidepressant drugs	1.231	(1.082–1.402)	0.0016
Antiplatelet drugs	1.187	(1.099–1.282)	<0.0001
Beta-adrenoceptor blocking drugs	0.899	(0.844–0.958)	0.0011
Bronchodilators	1.305	(1.174–1.452)	<0.0001
Drugs acting on the oropharynx	0.793	(0.729–0.863)	<0.0001
Drugs used in diabetes	1.753	(1.638–1.875)	<0.0001
Drugs used in parkinsonism and related disorders	2.255	(1.869–2.721)	<0.0001
Drugs used in psychoses and related disorders	1.369	(1.153–1.626)	0.0003
Laxatives	1.275	(1.186–1.37)	<0.0001
Miscellaneous drugs (nutrition and blood)	5.967	(2.432–14.638)	<0.0001
Minerals	1.161	(1.055–1.278)	0.0023
Positive inotropic drugs	1.504	(1.224–1.848)	0.0001
Vitamins	1.132	(1.051–1.22)	0.0011

CI: Confidence Interval; OR: Odds Ratio. Predictors in bold were among the top 20 predictors in both female and male models.

Table 5: **The top 20 predictors selected by stepwise selection by logistic regression models with the strongest association with hip fracture. The top 20 predictors selected by stepwise selection by logistic regression model for female with the strongest association with hip fracture.**

predicting the major osteoporotic fracture. With the genetic risk score, BMD and other known risk factors as predictors, they developed a prediction model with AUC of 0.71.[8] Since BMD, genotyping data and thus genetic risk score are not readily available among the public, this model also has limited generalizability. Another study used the administrative claims data of 288,086 individuals in Germany to develop an osteoporotic hip fracture prediction model with 4-year follow-up. Age, sex, history of fracture and medications known to be related to bone health were adopted as the predictors, attaining an AUC of 0.65–0.7.[9] Compared to these ML studies, our current study had sufficient sample size and the longest follow-up of 10 years. Notably, some of our ML models still had good discrimination performance (AUC > 0.8) even in the absence of BMD data. One plausible reason is the inclusion of all diagnosis and drug prescription records as potential predictors, as some comorbidities and drug use also contribute to BMD variation. This aligns with a previous proposal by the developers of fracture risk evaluation model (FREM) that the optimal prediction model should include both common (with known small or modest effects on fracture risk) and rare (whose relationship with fracture risk is yet to be revealed) risk factors.[27] The FREM utilized all the ICD-10 codes available from the Danish national register (n = 2,495,339) and applied backward selection by LR to develop one-year sex-stratified prediction models of hip fracture, attaining AUC of 0.87 and 0.85 for female and male respectively.[27] The inclusion of drug prescription records in our models may contribute to the good discrimination performance despite the smaller sample size. More importantly, the best-performing models for both female and male in the current study were the stepwise selection by LR models, but not the ML models. This is in line with a systematic review reporting that ML algorithms did not necessarily have better performance than LR model in clinical risk prediction, despite the flexibility of including nonlinear association and interaction terms in the model.[28]

A number of conventional risk factors were selected as the top 20 predictors by the LR models (Tables 5 and 6), such as age,[29] diagnosis and/or prescription records of accidental falls,[30] CVD,[31] chronic obstructive pulmonary diseases,[25] Parkinson's diseases,[32] epilepsy,[33] depression,[25] diabetes,[34] psychoses,[35] and nutritional deficiencies.[36] More importantly, our approach enables the identification of some relatively novel predictors of hip fracture. An example is drug prescription for anaemia and blood disorders, which was associated with higher odds of hip fracture (Tables 5 and 6). This is consistent with our recent Mendelian randomization

Predictors	Training cohort		
	OR	(95% CI)	p-value
Age on index date	1.164	(1.157–1.17)	<0.0001
Diagnosis			
Chronic obstructive pulmonary disease and allied conditions	1.63	(1.243–2.138)	0.0004
Organic psychotic conditions	2.153	(1.415–3.277)	0.0003
Other accidents	1.875	(1.315–2.673)	0.0005
Poisoning by drugs, medicinal and biological substances	10.166	(3.283–31.463)	<0.0001
Drug prescription			
Anaemias and some other blood disorders	1.608	(1.266–2.042)	<0.0001
Antiepileptic drugs	1.647	(1.277–2.124)	0.0001
Antiplatelet drugs	1.437	(1.295–1.594)	<0.0001
Beta-adrenoceptor blocking drugs	0.841	(0.763–0.926)	0.0004
Bronchodilators	1.547	(1.359–1.761)	<0.0001
Drugs acting on the oropharynx	0.804	(0.708–0.913)	0.0008
Drugs used in diabetes	1.431	(1.288–1.59)	<0.0001
Drugs used in parkinsonism and related disorders	3.209	(2.496–4.13)	<0.0001
Drugs used in psychoses and related disorders	1.998	(1.552–2.571)	<0.0001
Emollient and barrier preparations	1.448	(1.301–1.611)	<0.0001
Fluids and electrolytes	1.307	(1.134–1.507)	0.0002
Laxatives	1.373	(1.237–1.524)	<0.0001
Lipid-regulating drugs	0.744	(0.648–0.855)	<0.0001
Miscellaneous drugs (skin)	5.366	(1.981–14.52)	0.0009
Vitamins	1.399	(1.245–1.573)	<0.0001

CI: Confidence Interval; OR: Odds Ratio. Predictors in bold were among the top 20 predictors in both female and male models.

Table 6: The top 20 predictors selected by stepwise selection by logistic regression models with the strongest association with hip fracture. The top 20 predictors selected by stepwise selection by logistic regression model for male with the strongest association with hip fracture.

study that genetically determined red blood cell traits had positive causal effects on BMD.[37] Individuals with blood disorders, such as anaemia, may have lifelong risk of osteoporosis and fracture. In general, vitamins, laxatives, emollient are prescribed for poor appetite, constipation, and dry skin respectively. Together with anaemia, they are signs of ageing or frailty, which are the most important risk factor for fracture. Nevertheless, the exact underlying mechanisms of how the novel predictors might influence bone health or hip fracture warrant future investigations. On the other hand, some predictors were sex-specific, probably attributed to their different prevalence between sexes. An example is the diagnosis of nephritis, nephrotic syndrome and nephrosis, which was included in the female prediction model (Table 5). While chronic renal disease was adopted by QFracture as a risk factor irrespective of sex,[25] its related diagnosis was identified as a female-specific risk factor in our study, which partially aligned with previous literature that hip fracture incidence among women with chronic kidney diseases was twice as high as that in men.[38] Unexpectedly, history of fracture was not among the top 20 predictors in the LR model. Fracture of lower limb (including hip fracture) was ranked 27th and 32nd in the female and male prediction models, with odds ratio of 1.461 (95% CI:

1.125–1.897) and 2.02 (95% CI: 1.208–3.38) respectively. In addition, fracture of upper limb was also ranked 29th in the male prediction model, with odds ratio of 2.364 (95% CI: 1.295–4.314). One possible explanation for the low ranking of previous fracture in the LR model is the classification of disease status based on the ICD-9-CM sub-chapters. In our models, previous fracture is represented by the binary coding of four ICD-9-CM sub-chapters, including fracture of skull, fracture of neck and trunk, fracture of upper limb and fracture of lower limb. This may have diluted the significance of previous fracture as one single risk factor. Nevertheless, the odds ratio of previous fracture is consistently above 1 and it is the effect size which affects the calculated risk. Previous fracture still plays a key role in the LR model.

This study has several strengths and may be clinically important. We developed sex-specific hip fracture prediction models without utilizing clinical measurement data, such as BMD and body mass index (BMI). Yet, the best-performing prediction models have good discrimination performance with AUC >0.8. The female model also has adequate calibration. Using EHR data as the only predictors enables the integration of the prediction models into routine clinical workflow in the public healthcare setting. Amid the COVID-19 pandemic, healthcare services and resources were diverted to

combat COVID-19 and its related comorbidities from chronic diseases like osteoporosis.[39] Moreover, the prediction models were independently validated in a community-dwelling cohort. Taken together, despite the limited resources, the hip fracture prediction models may be applied at both public healthcare service setting and the public at population-level, aiding to triage individuals who are at high risk of hip fracture for prioritized DXA scan, and subsequent treatment initiation. Such measures are expected to facilitate early prevention, timely diagnosis and treatment of osteoporosis.

Our study also has limitations. First, diagnosis and prescription records within one year prior to the index date were retrieved in the current study. Yet, the diagnosis of chronic diseases might not be repeatedly coded in CDARS, explaining why the top 20 predictors were mainly drug prescription variables. Notably, medication use was recorded in CDARS upon prescription regardless of the onset of the disease. Thus, the inclusion of drug prescription variables is complementary to the use of diagnosis variables. Future incorporation of laboratory test results in the prediction models may further improve the model performance. Second, the electronic medical database did not capture risk factors related to lifestyle (such as alcohol consumption and smoking) and clinical measurement (such as BMI and weight). Nevertheless, these may be proxied by the diagnosis and drug prescription records available. Third, although CDARS data is representative and 98% of hip fracture cases in Hong Kong was admitted to HA hospitals, loss to follow-up due to emigration and death is inevitable. While emigration could be considered random, people with longer life expectancy are more likely to have complete follow-up, whom the models may have more favourable predictive performance. Fourth, the generalizability of the model to other populations is unclear. External validation cohorts in other populations may not be linked to any electronic records, or they may be linked to claims database instead of EHR. Even if the external cohorts were linked to EHR, different coding systems like ICD-10-CM or read codes (used by The Health Improvement Network [THIN]) may be adopted. As accurate conversion of the diagnosis and drug prescription records to ICD-9-CM and BNF coding are required respectively, it is particularly challenging to validate the prediction models in external validation cohorts. In addition to the intrinsic difference among populations, such as ethnicities, demographics and lifestyle factors, further evaluation of the prediction models in external validation cohorts are warranted to determine its generalizability.

In conclusion, we have developed and validated sex-specific hip fracture prediction tools at population-level in Hong Kong using EHR. Notably, the good discrimination and calibration performance of the LR model for female was validated in both internal and independent cohorts, implying that the model may be clinically useful and generalizable to the public. Despite the high discrimination performance, the LR model for male would require additional calibration in independent cohorts. By using EHR as predictors, it is expected that the prediction model could be integrated to the routine clinical workflow, assisting clinicians to identify people who are at high risk of hip fracture for DXA scan. These measures may facilitate early prevention, timely diagnosis and treatment of osteoporosis.

Appendix A. Supplementary data
Supplementary data related to this article can be found at https://doi.org/10.1016/j.eclinm.2023.101876.

References
1 Cooper C, Campion G, Melton LJ 3rd. Hip fractures in the elderly: a world-wide projection. *Osteoporos Int.* 1992;2(6):285–289.
2 Gullberg B, Johnell O, Kanis JA. World-wide projections for hip fracture. *Osteoporos Int.* 1997;7(5):407–413.
3 Cheung CL, Ang SB, Chadha M, et al. An updated hip fracture projection in Asia: the Asian Federation of Osteoporosis Societies study. *Osteoporos Sarcopenia.* 2018;4(1):16–21.

4   Handa R, Ali Kalla A, Maalouf G. Osteoporosis in developing countries. *Best Pract Res Clin Rheumatol*. 2008;22(4):693–708.

5   Kanis JA, Norton N, Harvey NC, et al. Scope 2021: a new scorecard for osteoporosis in Europe. *Arch Osteoporos*. 2021;16(1):82.

6   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;19(4):385–397.

7   Cheung EY, Bow CH, Cheung CL, et al. Discriminative value of FRAX for fracture prediction in a cohort of Chinese post-menopausal women. *Osteoporos Int*. 2012;23(3):871–878.

8   Wu Q, Nasoz F, Jung J, Bhattarai B, Han MV. Machine learning approaches for fracture risk assessment: a comparative analysis of genomic and phenotypic data in 5130 older men. *Calcif Tissue Int*. 2020;107(4):353–361.

9   Engels A, Reber KC, Lindlbauer I, et al. Osteoporotic hip fracture prediction from risk factors available in administrative claims data - a machine learning approach. *PLoS One*. 2020;15(5):e0232969.

10   Kruse C, Eiken P, Vestergaard P. Machine learning principles can improve hip fracture prediction. *Calcif Tissue Int*. 2017;100(4):348–360.

11   The Hong Kong Hospital Authority. The Hong Kong Hospital Authority statisitical report 2016-2017. https://www3.ha.org.hk/data/HAStatistics/DownloadReport/2; 2017. Accessed June 13, 2022.

12   Sing CW, Woo YC, Lee ACH, et al. Validity of major osteoporotic fracture diagnosis codes in the clinical data analysis and reporting system in Hong Kong. *Pharmacoepidemiol Drug Saf*. 2017;26(8):973–976.

13   Cheung CL, Tan KCB, Kung AWC. Cohort profile: the Hong Kong osteoporosis study and the follow-up study. *Int J Epidemiol*. 2017;47(2):397–398f.

14   Wasey JO. Icd - fast comorbidities from ICD-9 and ICD-10 codes, decoding, manipulation and validation. https://www.rdocumentation.org/packages/icd/versions/4.0.9; 2020. Accessed June 15, 2022.

15   Bruce A, Bruce P. *Regression and prediction. Practical statistics for data scientists*. 1st ed. O'Reilly Media, Inc.; 2017.

16   Ripley B. MASS: support functions and datasets for venables and ripley's MASS. https://cran.r-project.org/web/packages/MASS/index.html; 2019. Accessed November 12, 2019.

17   Kuhn M. The caret package. https://topepo.github.io/caret/; 2019. Accessed June 15, 2022.

18   Perkins NJ, Schisterman EF. The inconsistency of "optimal" cut-points obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol*. 2006;163(7):670–675.

19   Harrell FE. Package "Hmisc". https://cran.r-project.org/web/packages/Hmisc/index.html; 2019.

20   Huang Y, Li W, Macheret F, Gabriel RA, Ohno-Machado L. A tutorial on calibration measurements and calibration models for clinical prediction models. *J Am Med Inform Assoc*. 2020;27(4):621–633.

21   Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Stat Med*. 1999;18(17–18):2529–2545.

22   Spiegelhalter DJ. Probabilistic prediction in patient management and clinical trials. *Stat Med*. 1986;5(5):421–433.

23   Harrell FE. Package rms: regression modeling strategies. https://cran.r-project.org/web/packages/rms/index.html; 2022. Accessed July 14, 2022.

24   Schummers L, Himes KP, Bodnar LM, Hutcheon JA. Predictor characteristics necessary for building a clinically useful risk prediction model: a simulation study. *BMC Med Res Methodol*. 2016;16(1):123.

25   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;344:e3427.

26   Nguyen ND, Frost SA, Center JR, Eisman JA, Nguyen TV. Development of prognostic nomograms for individualizing 5-year and 10-year fracture risks. *Osteoporos Int*. 2008;19(10):1431–1444.

27   Rubin KH, Moller S, Holmberg T, Bliddal M, Sondergaard J, Abrahamsen B. A new fracture risk assessment tool (FREM) based on public health registries. *J Bone Miner Res*. 2018;33(11):1967–1979.

28   Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12–22.

29   Kanis JA, Oden A, Johnell O, et al. The use of clinical risk factors enhances the performance of BMD in the prediction of hip and osteoporotic fractures in men and women. *Osteoporos Int*. 2007;18(8):1033–1046.

30   Nachreiner NM, Findorff MJ, Wyman JF, McCarthy TC. Circumstances and consequences of falls in community-dwelling older women. *J Womens Health (Larchmt)*. 2007;16(10):1437–1446.

31   Sennerby U, Melhus H, Gedeborg R, et al. Cardiovascular diseases and risk of hip fracture. *JAMA*. 2009;302(15):1666–1673.

32   Chen YY, Cheng PY, Wu SL, Lai CH. Parkinson's disease and risk of hip fracture: an 8-year follow-up study in Taiwan. *Parkinsonism Relat Disord*. 2012;18(5):506–509.

33   Jette N, Lix LM, Metge CJ, Prior HJ, McChesney J, Leslie WD. Association of antiepileptic drugs with nontraumatic fractures: a population-based analysis. *Arch Neurol*. 2011;68(1):107–112.

34   Robbins J, Aragaki AK, Kooperberg C, et al. Factors associated with 5-year risk of hip fracture in postmenopausal women. *JAMA*. 2007;298(20):2389–2398.

35   Takkouche B, Montes-Martinez A, Gill SS, Etminan M. Psycho-tropic medications and the risk of fracture: a meta-analysis. *Drug Saf*. 2007;30(2):171–184.

36   Gennari C. Calcium and vitamin D nutrition and bone disease of the elderly. *Public Health Nutr*. 2001;4(2B):547–559.

37   Ho SC, Li GH, Leung AY, Tan KC, Cheung CL. Unravelling genetic causality of haematopoiesis on bone metabolism in human. *Eur J Endocrinol*. 2022;187(6):765–775.

38   Pimentel A, Urena-Torres P, Zillikens MC, Bover J, Cohen-Solal M. Fractures in patients with CKD-diagnosis, treatment, and prevention: a review by members of the European Calcified Tissue Society and the European Renal Association of Nephrology Dialysis and Transplantation. *Kidney Int*. 2017;92(6):1343–1355.

39   Fuggle NR, Singer A, Gill C, et al. How has COVID-19 affected the treatment of osteoporosis? An IOF-NOF-ESCEO global survey. *Osteoporos Int*. 2021;32(4):611–617.