Machine learning models for hip fracture prediction using electronic medical records: abridged secondary publication

GHY Li *, CL Cheung, KCB Tan, TCY Kwok, WCY Lau

Hong Kong Med J 2025;31(Suppl 3):S4-11
HMRF project number: 17181381

[1] GHY Li, [2] CL Cheung, [3] KCB Tan, [4] TCY Kwok, [5] WCY Lau

[1] *Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong SAR, China*
[2] *Department of Pharmacology and Pharmacy, The University of Hong Kong, Hong Kong SAR, China*
[3] *Department of Medicine, The University of Hong Kong, Hong Kong SAR, China*
[4] *Department of Medicine & Therapeutics and School of Public Health, The Chinese University of Hong Kong, Hong Kong SAR, China*
[5] *School of Pharmacy, University College London, United Kingdom*

* Principal applicant and corresponding author: gloria-hy.li@polyu.edu.hk

KEY MESSAGES

1. Ethnicity- and sex-specific hip fracture prediction models were developed using machine learning algorithms and electronic medical records. The performance of the prediction models was validated in independent cohorts, achieving the area under the curve values of >0.8. The prediction models may be clinically useful and generalisable to the public.
2. The prediction models were developed without using bone mineral density as a potential predictor, owing to the limited availability of dual-energy X-ray absorptiometry in Hong Kong.

Introduction

Osteoporosis is characterised by low bone mass and deterioration in bone strength and microarchitecture, increasing the risk of fragility fracture. In 2050, the number of hip fractures in Asia is expected to reach 2.56 million.[1] In Hong Kong, an estimated 27 468 hip fractures will occur in that year, costing approximately HK$1.9 billion annually.[1]

Fracture prediction models, such as the Fracture Risk Assessment Tool, are mainly developed using data from Western populations. Among the Hong Kong population, fracture prediction based on ethnicity-specific clinical risk factors and femoral neck bone mineral density (BMD) T-score outperforms that of the Fracture Risk Assessment Tool.[2] Dual-energy X-ray absorptiometry is the gold standard in diagnosing osteoporosis and predicting fracture. However, its availability is considerably low. In Hong Kong public hospitals, the median waiting time for a scan is 9 months.[3] Therefore, it is important to develop a fracture risk prediction tool without BMD data. This study aimed to develop and validate sex- and ethnicity-specific machine learning models to predict the 10- and 15-year hip fracture risks, based on demographic, diagnostic, and prescription data from electronic medical records.

Methods

Anonymised medical records were retrieved from the Clinical Data Analysis and Reporting System of the Hospital Authority. As of 31 December 2005 (index date), approximately 740 000 public healthcare service users aged ≥60 years had admission records between 1 January and 31 December 2005. Around one-third of them were randomly selected as the derivation cohort; they were followed up until hip fracture, death, or study end dates (31 December 2015 and 31 December 2020 for 10- and 15-year risk prediction, respectively), whichever occurred earlier. The derivation cohort was stratified by sex, and each sex-specific sub-cohort was randomly divided into an internal training (80%) and internal testing (20%) dataset. Performance of the prediction models was further assessed in an external validation cohort comprising participants aged ≥60 years from the Hong Kong Osteoporosis Study,[4] which comprised 9449 community-dwelling Southern Chinese participants. The external validation cohort did not overlap with the derivation cohort. In the supplementary analysis, individuals who had been lost to follow-up or died before study end dates were excluded.

Potential predictors included age, number of hospitalisations, and diagnosis and drug prescription records within 1 year of the index date. The presence or absence of each diagnosis code and drug class was recorded. Initially, there were 396 potential predictors (162 diagnosis variables and 232 drug prescription variables). Variables with zero or near-zero variance (≤0.1% prevalence) were excluded.

A multistep model selection procedure for logistic regression (LR) was applied to the development dataset (ie, 80% of the derivation cohort). This approach to finding the model that penalised variable addition involved minimising the metric of the Akaike information criterion (ie, including all predictors at the start, then performing stepwise selection by LR, which added and dropped predictors to identify a model with the lowest Akaike information criterion).

TABLE 1a. Characteristics of female participants.

Variable	10-year risk*						15-year risk*					
	Whole sample			Excluding those lost to follow-up			Whole sample			Excluding those lost to follow-up		
	Derivation cohort		External validation cohort (n=2520)	Derivation cohort		External validation cohort (n=2038)	Derivation cohort		External validation cohort (n=2520)	Derivation cohort		External validation cohort (n=1762)
	Training cohort (n=103515)	Testing cohort (n=25878)		Training cohort (n=73541)	Testing cohort (n=18385)		Training cohort (n=103515)	Testing cohort (n=25878)		Training cohort (n=60659)	Testing cohort (n=15164)	
Hip fracture cases	7568 (7.3)	1892 (7.3)	145 (5.8)‡	7568 (10.3)	1892 (10.3)	145 (7.1)‡	10743 (10.4)	2685 (10.4)	211 (8.4)‡	10743 (17.7)	2685 (17.7)	211 (12.0)‡
Age, y	73 (60-114)	73 (60-114)	71 (60-100)‡	71 (60-106)	71 (60-103)	69 (60-95)‡	73 (60-114)	73 (60-107)	71 (60-100)‡	70 (60-106)	70 (60-100)	68 (60-95)‡
No. of hospital admissions	0 (0-22)	0 (0-21)	0 (0-9)‡	0 (0-21)	0 (0-21)	0 (0-9)‡	0 (0-21)	0 (0-22)	0 (0-9)‡	0 (0-13)	0 (0-21)	0 (0-9)‡
Diagnosis (within 1 year)												
Chronic obstructive pulmonary disease and allied conditions	1557 (1.5)	400 (1.5)	21 (0.8)	574 (0.8)	130 (0.7)	10 (0.5)	1573 (1.5)	384 (1.5)	21 (0.8)‡	408 (0.7)	105 (0.7)	7 (0.4)
Any cancer												
Malignant neoplasm of lip, oral cavity, and pharynx	42 (0.0)	5 (0.0)	0 (0.0)	15 (0.0)	4 (0.0)	0 (0.0)	35 (0.0)	12 (0.0)	0 (0.0)	9 (0.0)	3 (0.0)	0 (0.0)
Malignant neoplasm of digestive organs and peritoneum	401 (0.4)	96 (0.4)	7 (0.3)	140 (0.2)	40 (0.2)	2 (0.1)	402 (0.4)	95 (0.4)	7 (0.3)	131 (0.2)	22 (0.1)	2 (0.1)
Malignant neoplasm of respiratory and intrathoracic organs	157 (0.2)	34 (0.1)	4 (0.2)	22 (0.0)	3 (0.0)	0 (0.0)	143 (0.1)	48 (0.2)	4 (0.2)	18 (0.0)	5 (0.0)	0 (0.0)
Malignant neoplasm of bone, connective tissue, skin, and breast	257 (0.2)	65 (0.3)	8 (0.3)	135 (0.2)	36 (0.2)	6 (0.3)	257 (0.2)	65 (0.3)	8 (0.3)	111 (0.2)	28 (0.2)	6 (0.3)
Malignant neoplasm of genitourinary organs	228 (0.2)	60 (0.2)	2 (0.1)	98 (0.1)	30 (0.2)	2 (0.1)	241 (0.2)	47 (0.2)	2 (0.1)	75 (0.1)	24 (0.2)	2 (0.1)
Malignant neoplasm of other and unspecified sites	316 (0.3)	76 (0.3)	3 (0.1)	55 (0.1)	20 (0.1)	1 (0.0)	322 (0.3)	70 (0.3)	3 (0.1)	49 (0.1)	12 (0.1)	1 (0.1)
Malignant neoplasm of lymphatic and haematopoietic tissue	94 (0.1)	17 (0.1)	6 (0.2)‡	18 (0.0)	5 (0.0)	3 (0.1)‡	94 (0.1)	17 (0.1)	6 (0.2)‡	13 (0.0)	6 (0.0)	2 (0.1)
Cardiovascular disease												
Chronic rheumatic heart disease	193 (0.2)	31 (0.1)†	1 (0.0)	64 (0.1)	23 (0.1)	1 (0.0)	189 (0.2)	35 (0.1)	1 (0.0)	49 (0.1)	14 (0.1)	1 (0.1)
Hypertensive disease	4319 (4.2)	1098 (4.2)	87 (3.5)	1937 (2.6)	497 (2.7)	48 (2.4)	4328 (4.2)	1089 (4.2)	87 (3.5)	1462 (2.4)	301 (2.0)†	38 (2.2)
Ischaemic heart disease	1898 (1.8)	451 (1.7)	31 (1.2)‡	780 (1.1)	230 (1.3)†	17 (0.8)	1884 (1.8)	465 (1.8)	31 (1.2)‡	604 (1.0)	139 (0.9)	13 (0.7)
Diseases of pulmonary circulation	57 (0.1)	7 (0.0)	1 (0.0)	18 (0.0)	3 (0.0)	0 (0.0)	52 (0.1)	12 (0.0)	1 (0.0)	18 (0.0)	2 (0.0)	0 (0.0)
Other forms of heart disease	3102 (3.0)	777 (3.0)	40 (1.6)‡	952 (1.3)	232 (1.3)	15 (0.7)‡	3091 (3.0)	788 (3.0)	40 (1.6)‡	674 (1.1)	193 (1.3)	11 (0.6)
Cerebrovascular disease	2261 (2.2)	576 (2.2)	35 (1.4)‡	829 (1.1)	217 (1.2)	12 (0.6)‡	2254 (2.2)	583 (2.3)	35 (1.4)‡	608 (1.0)	157 (1.0)	7 (0.4)‡
Diseases of arteries, arterioles, and capillaries	366 (0.4)	95 (0.4)	11 (0.4)	130 (0.2)	25 (0.1)	4 (0.2)	371 (0.4)	90 (0.3)	11 (0.4)	86 (0.1)	30 (0.2)	2 (0.1)
Diseases of veins and lymphatics, and other diseases of circulatory system	767 (0.7)	163 (0.6)	18 (0.7)	383 (0.5)	97 (0.5)	13 (0.6)	750 (0.7)	180 (0.7)	18 (0.7)	308 (0.5)	71 (0.5)	10 (0.6)
Psychotic conditions, including dementias and alcohol-induced mental disorders	1565 (1.5)	367 (1.4)	28 (1.1)	374 (0.5)	101 (0.5)	4 (0.2)	1547 (1.5)	385 (1.5)	28 (1.1)	272 (0.4)	53 (0.3)	4 (0.2)
Rheumatism, excluding the back	438 (0.4)	113 (0.4)	14 (0.6)	296 (0.4)	86 (0.5)	11 (0.5)	436 (0.4)	115 (0.4)	14 (0.6)	254 (0.4)	69 (0.5)	10 (0.6)
Nephritis, nephrotic syndrome, and nephrosis	934 (0.9)	212 (0.8)	15 (0.6)	170 (0.2)	44 (0.2)	2 (0.1)	917 (0.9)	229 (0.9)	15 (0.6)	119 (0.2)	38 (0.3)	1 (0.1)
Diseases of endocrine glands, including diabetes mellitus, disorders of pituitary and parathyroid glands, and ovarian dysfunction	3230 (3.1)	796 (3.1)	57 (2.3)‡	1290 (1.8)	353 (1.9)	25 (1.2)	3218 (3.1)	808 (3.1)	57 (2.3)‡	919 (1.5)	242 (1.6)	17 (1.0)
Previous fracture												
Fracture of skull	46 (0.0)	9 (0.0)	2 (0.1)	23 (0.0)	9 (0.0)	2 (0.1)	43 (0.0)	12 (0.0)	2 (0.1)	22 (0.0)	2 (0.0)	2 (0.1)
Fracture of neck and trunk	395 (0.4)	86 (0.3)	14 (0.6)	188 (0.3)	54 (0.3)	6 (0.3)	372 (0.4)	109 (0.4)	14 (0.6)	141 (0.2)	40 (0.3)	5 (0.3)
Fracture of upper limb	879 (0.8)	201 (0.8)	22 (0.9)	496 (0.7)	123 (0.7)	15 (0.7)	882 (0.9)	198 (0.8)	22 (0.9)	413 (0.7)	91 (0.6)	13 (0.7)
Fracture of lower limb	1236 (1.2)	324 (1.3)	47 (1.9)‡	599 (0.8)	144 (0.8)	23 (1.1)	1229 (1.2)	331 (1.3)	47 (1.9)‡	486 (0.8)	107 (0.7)	17 (1.0)
Drug prescription (within 1 year)												
Any antidepressants (BNF 4.3)	5019 (4.8)	1236 (4.8)	88 (3.5)‡	2856 (3.9)	764 (4.2)	62 (3.0)‡	5015 (4.8)	1240 (4.8)	88 (3.5)‡	2270 (3.7)	559 (3.7)	56 (3.2)
Drugs used in rheumatic diseases and gout (BNF 10.1)	19080 (18.4)	4719 (18.2)	336 (13.3)‡	13516 (18.4)	3380 (18.4)	270 (13.2)‡	19050 (18.4)	4749 (18.4)	336 (13.3)‡	11090 (18.3)	2878 (19.0)	235 (13.3)‡
Corticosteroids												
Respiratory (BNF 3.2)	2488 (2.4)	646 (2.5)	40 (1.6)‡	1294 (1.8)	281 (1.5)†	27 (1.3)	2562 (2.5)	572 (2.2)	40 (1.6)‡	976 (1.6)	220 (1.5)	22 (1.2)
Endocrine (BNF 6.3)	3845 (3.7)	962 (3.7)	51 (2.0)‡	1875 (2.5)	484 (2.6)	25 (1.2)‡	3887 (3.8)	920 (3.6)	51 (2.0)‡	1462 (2.4)	355 (2.3)	20 (1.1)‡
Topical (BNF 13.4)	14975 (14.5)	3785 (14.6)	212 (8.4)‡	10165 (13.8)	2454 (13.3)	175 (8.6)‡	15031 (14.5)	3729 (14.4)	212 (8.4)‡	8167 (13.5)	2061 (13.6)	145 (8.2)‡

Abbreviation: BNF=British National Formulary

* Data are presented as median (range) or No. (%) of participants.

† P<0.05 between training and testing cohorts

‡ P<0.05 between derivation and external validation cohorts

TABLE 1b. Characteristics of male participants.

Variable	10-year risk*						15-year risk*					
	Whole sample			Excluding those lost to follow-up			Whole sample			Excluding those lost to follow-up		
	Derivation cohort		External validation cohort (n=1277)	Derivation cohort		External validation cohort (n=1008)	Derivation cohort		External validation cohort (n=1277)	Derivation cohort		External validation cohort (n=834)
	Training cohort (n=88 483)	Testing cohort (n=22 120)		Training cohort (n=55 301)	Testing cohort (n=13 824)		Training cohort (n=88 483)	Testing cohort (n=22 120)		Training cohort (n=41 877)	Testing cohort (n=10 468)	
Hip fracture cases	3301 (3.7)	825 (3.7)	36 (2.8)	3301 (6.0)	825 (6.0)	36 (3.6)‡	4791 (5.4)	1197 (5.4)	58 (4.5)	4791 (11.4)	1197 (11.4)	58 (7.0)‡
Age, y	71 (60-107)	71 (60-105)	70 (60-96)‡	69 (60-104)	69 (60-101)	68 (60-96)	71 (60-107)	71 (60-105)	70 (60-96)‡	68 (60-104)	68 (60-103)	68 (60-96)
No. of hospital admissions	0 (0-42)	0 (0-24)	0 (0-8)‡	0 (0-42)	0 (0-18)	0 (0-4)‡	0 (0-30)	0 (0-42)	0 (0-8)‡	0 (0-18)	0 (0-21)	0 (0-3)‡
Diagnosis (within 1 year)												
Chronic obstructive pulmonary disease and allied conditions	2927 (3.3)	729 (3.3)	29 (2.3)‡	716 (1.3)	174 (1.3)	8 (0.8)	2912 (3.3)	744 (3.4)	29 (2.3)‡	482 (1.2)	109 (1.0)	5 (0.6)
Any cancer												
Malignant neoplasm of lip, oral cavity, and pharynx	89 (0.1)	22 (0.1)	1 (0.1)	31 (0.1)	4 (0.0)	1 (0.1)	82 (0.1)	29 (0.1)	1 (0.1)	20 (0.0)	6 (0.1)	0 (0.0)
Malignant neoplasm of digestive organs and peritoneum	700 (0.8)	180 (0.8)	5 (0.4)	186 (0.3)	46 (0.3)	3 (0.3)	708 (0.8)	172 (0.8)	5 (0.4)	130 (0.3)	39 (0.4)	2 (0.2)
Malignant neoplasm of respiratory and intrathoracic organs	320 (0.4)	88 (0.4)	1 (0.1)	62 (0.1)	15 (0.1)	0 (0.0)	326 (0.4)	82 (0.4)	1 (0.1)	45 (0.1)	13 (0.1)	0 (0.0)
Malignant neoplasm of bone, connective tissue, skin, and breast	44 (0.0)	14 (0.1)	3 (0.2)‡	21 (0.0)	6 (0.0)	0 (0.0)‡	45 (0.1)	13 (0.1)	3 (0.2)‡	15 (0.0)	5 (0.0)	3 (0.4)‡
Malignant neoplasm of genitourinary organs	613 (0.7)	169 (0.8)	3 (0.2)	217 (0.4)	66 (0.5)	2 (0.2)	601 (0.7)	181 (0.8)†	3 (0.2)	160 (0.4)	31 (0.3)	1 (0.1)
Malignant neoplasm of other and unspecified sites	449 (0.5)	113 (0.5)	1 (0.1)‡	69 (0.1)	18 (0.1)	1 (0.1)	437 (0.5)	125 (0.6)	1 (0.1)	60 (0.1)	14 (0.1)	0 (0.0)
Malignant neoplasm of lymphatic and haematopoietic tissue	95 (0.1)	40 (0.2)	1 (0.1)	26 (0.0)	9 (0.1)	0 (0.0)	106 (0.1)	29 (0.1)	1 (0.1)	19 (0.0)	5 (0.0)	0 (0.0)
Cardiovascular disease												
Chronic rheumatic heart disease	92 (0.1)	32 (0.1)	1 (0.1)	42 (0.1)	11 (0.1)	1 (0.1)	99 (0.1)	25 (0.1)	1 (0.1)	22 (0.1)	10 (0.1)	0 (0.0)
Hypertensive disease	3714 (4.2)	915 (4.1)	31 (2.4)‡	1439 (2.6)	374 (2.7)	15 (1.5)‡	3669 (4.1)	960 (4.3)	31 (2.4)‡	1002 (2.4)	251 (2.4)	8 (1.0)‡
Ischaemic heart disease	2248 (2.5)	557 (2.5)	23 (1.8)	990 (1.8)	240 (1.7)	7 (0.7)‡	2234 (2.5)	571 (2.6)	23 (1.8)	663 (1.6)	175 (1.7)	5 (0.6)‡
Diseases of pulmonary circulation	64 (0.1)	20 (0.1)	1 (0.1)	11 (0.0)	7 (0.1)	1 (0.1)	67 (0.1)	17 (0.1)	1 (0.1)	8 (0.0)	5 (0.0)	1 (0.1)
Other forms of heart disease	2532 (2.9)	659 (3.0)	29 (2.3)	756 (1.4)	211 (1.5)	9 (0.9)	2558 (2.9)	633 (2.9)	29 (2.3)	506 (1.2)	126 (1.2)	6 (0.7)
Cerebrovascular disease	2426 (2.7)	596 (2.7)	21 (1.6)‡	850 (1.5)	196 (1.4)	10 (1.0)	2408 (2.7)	614 (2.8)	21 (1.6)‡	572 (1.4)	139 (1.3)	8 (1.0)
Diseases of arteries, arterioles, and capillaries	461 (0.5)	113 (0.5)	6 (0.5)	151 (0.3)	27 (0.2)	2 (0.2)	455 (0.5)	119 (0.5)	6 (0.5)	91 (0.2)	26 (0.2)	0 (0.0)
Diseases of veins and lymphatics, and other diseases of circulatory system	825 (0.9)	200 (0.9)	8 (0.6)	380 (0.7)	108 (0.8)	6 (0.6)	819 (0.9)	206 (0.9)	8 (0.6)	300 (0.7)	79 (0.8)	4 (0.5)
Psychotic conditions, including dementias and alcohol-induced mental disorders	922 (1.0)	230 (1.0)	15 (1.2)	158 (0.3)	43 (0.3)	1 (0.1)	929 (1.0)	223 (1.0)	15 (1.2)	115 (0.3)	35 (0.3)	1 (0.1)
Rheumatism, excluding the back	307 (0.3)	76 (0.3)	4 (0.3)	183 (0.3)	46 (0.3)	3 (0.3)	303 (0.3)	80 (0.4)	4 (0.3)	135 (0.3)	35 (0.3)	3 (0.4)
Nephritis, nephrotic syndrome, and nephrosis	967 (1.1)	237 (1.1)	7 (0.5)	170 (0.3)	61 (0.4)†	3 (0.3)	954 (1.1)	250 (1.1)	7 (0.5)	135 (0.3)	30 (0.3)	2 (0.2)
Diseases of endocrine glands, including diabetes mellitus, disorders of pituitary and parathyroid glands, and ovarian dysfunction	2680 (3.0)	694 (3.1)	15 (1.2)‡	967 (1.7)	229 (1.7)	7 (0.7)‡	2730 (3.1)	644 (2.9)	15 (1.2)‡	618 (1.5)	165 (1.6)	5 (0.6)‡
Previous fracture												
Fracture of skull	53 (0.1)	10 (0.0)	0 (0.0)	20 (0.0)	6 (0.0)	0 (0.0)	49 (0.1)	14 (0.1)	0 (0.0)	20 (0.0)	4 (0.0)	0 (0.0)
Fracture of neck and trunk	139 (0.2)	26 (0.1)	0 (0.0)	68 (0.1)	18 (0.1)	0 (0.0)	137 (0.2)	28 (0.1)	0 (0.0)	51 (0.1)	11 (0.1)	0 (0.0)
Fracture of upper limb	251 (0.3)	58 (0.3)	2 (0.2)	117 (0.2)	42 (0.3)	2 (0.2)	243 (0.3)	66 (0.3)	2 (0.2)	99 (0.2)	26 (0.2)	0 (0.0)
Fracture of lower limb	496 (0.6)	116 (0.5)	3 (0.2)	209 (0.4)	59 (0.4)	2 (0.2)	483 (0.5)	129 (0.6)	3 (0.2)	185 (0.4)	33 (0.3)	1 (0.1)
Drug prescription (within 1 year)												
Any antidepressants (BNF 4.3)	2236 (2.5)	532 (2.4)	29 (2.3)	1117 (2.0)	248 (1.8)	15 (1.5)	2190 (2.5)	578 (2.6)	29 (2.3)	795 (1.9)	183 (1.7)	12 (1.4)
Drugs used in rheumatic diseases and gout (BNF 10.1)	16443 (18.6)	4213 (19.0)	161 (12.6)‡	9908 (17.9)	2511 (18.2)	124 (12.3)‡	16481 (18.6)	4175 (18.9)	161 (12.6)‡	7494 (17.9)	1841 (17.6)	94 (11.3)‡
Corticosteroids												
Respiratory (BNF 3.2)	3918 (4.4)	990 (4.5)	31 (2.4)‡	1346 (2.4)	307 (2.2)	13 (1.3)‡	3961 (4.5)	947 (4.3)	31 (2.4)‡	852 (2.0)	228 (2.2)	7 (0.8)‡
Endocrine (BNF 6.3)	5144 (5.8)	1335 (6.0)	36 (2.8)‡	1863 (3.4)	461 (3.3)	19 (1.9)‡	5198 (5.9)	1281 (5.8)	36 (2.8)‡	1314 (3.1)	307 (2.9)	12 (1.4)‡
Topical (BNF 13.4)	12363 (14.0)	3090 (14.0)	135 (10.6)‡	7175 (13.0)	1834 (13.3)	95 (9.4)‡	12341 (13.9)	3112 (14.1)	135 (10.6)‡	5347 (12.8)	1318 (12.6)	75 (9.0)‡

Abbreviation: BNF=British National Formulary

* Data are presented as median (range) or No. (%) of participants.

† P<0.05 between training and testing cohorts

‡ P<0.05 between derivation and external validation cohorts

TABLE 2a. Performance of hip fracture risk prediction models for women.

Performance	Stepwise selection by logistic regression		Gradient boosting machine		Random forest		Extreme gradient boosting		Neural networks with a single hidden layer		Naïve Bayes	
	10-year risk	15-year risk	10-year risk	15-year risk	10-year risk	15-year risk	10-year risk	15-year risk	10-year risk	15-year risk	10-year risk	15-year risk
Training cohort												
Area under the curve (95% confidence interval)	0.716 (0.71-0.721)	0.669 (0.664-0.674)	0.687 (0.682-0.693)	0.643 (0.638-0.648)	0.997 (0.996-0.997)	0.996 (0.996-0.996)	0.675 (0.669-0.681)	0.631 (0.626-0.636)	0.691 (0.685-0.697)	0.66 (0.655-0.665)	0.521 (0.514-0.528)	0.519 (0.513-0.525)
Testing cohort												
Area under the curve (95% confidence interval)	0.705 (0.694-0.716)	0.652 (0.641-0.662)	0.677 (0.665-0.689)	0.627 (0.616-0.637)	0.65 (0.638-0.661)	0.593 (0.582-0.603)	0.663 (0.651-0.675)	0.61 (0.599-0.621)	0.67 (0.658-0.683)	0.623 (0.612-0.634)	0.5 (0.5-0.5)	0.478 (0.469-0.487)
Sensitivity	0.724	0.687	0.715	0.694	0.11	0.102	0.721	0.591	0.687	0.587	0.244	0.202
Specificity	0.59	0.542	0.55	0.51	0.929	0.924	0.526	0.581	0.586	0.59	0.817	0.844
Positive predictive value	0.122	0.148	0.111	0.141	0.109	0.135	0.107	0.14	0.116	0.142	0.095	0.131
Negative predictive value	0.964	0.937	0.961	0.935	0.93	0.899	0.96	0.925	0.96	0.925	0.932	0.901
F1	0.209	0.244	0.193	0.234	0.109	0.116	0.187	0.227	0.198	0.229	0.137	0.159
Accuracy	0.6	0.557	0.562	0.53	0.869	0.839	0.54	0.582	0.594	0.59	0.775	0.778
Error	0.4	0.443	0.438	0.471	0.131	0.161	0.46	0.418	0.406	0.41	0.225	0.222
External validation cohort												
Area under the curve (95% confidence interval)	0.769 (0.734-0.803)	0.724 (0.691-0.757)	0.757 (0.718-0.795)	0.702 (0.667-0.738)	0.669 (0.629-0.708)	0.607 (0.569-0.645)	0.745 (0.705-0.784)	0.691 (0.653-0.728)	0.713 (0.67-0.757)	0.685 (0.651-0.72)	0.5 (0.5-0.5)	0.456 (0.419-0.494)
Sensitivity	0.724	0.697	0.772	0.758	0.241	0.242	0.807	0.697	0.717	0.645	0.414	0.341
Specificity	0.699	0.661	0.621	0.573	0.852	0.849	0.581	0.615	0.648	0.64	0.726	0.743
Positive predictive value	0.128	0.158	0.111	0.14	0.091	0.128	0.105	0.142	0.111	0.141	0.085	0.108
Negative predictive value	0.976	0.96	0.978	0.963	0.949	0.924	0.98	0.957	0.974	0.952	0.953	0.925
F1	0.218	0.258	0.194	0.236	0.132	0.167	0.186	0.236	0.192	0.231	0.14	0.164
Accuracy	0.7	0.664	0.63	0.589	0.817	0.798	0.594	0.621	0.652	0.641	0.708	0.71
Error	0.3	0.336	0.37	0.411	0.183	0.202	0.406	0.379	0.348	0.36	0.292	0.291
Delong's test P value	Reference	Reference	0.364	0.12	<0.001	<0.001	0.09	0.032	<0.001	0.005	<0.001	<0.001
Integrated discrimination improvement (P value)	Reference	Reference	-0.174 (<0.001)	-0.117 (<0.001)	-0.133 (<0.001)	-0.099 (<0.001)	-0.185 (<0.001)	-0.129 (<0.001)	-0.196 (<0.001)	-0.159 (<0.001)	-0.04 (<0.001)	-0.037 (<0.001)

Five machine learning algorithms were used to train the prediction model: gradient boosting machine, random forest, extreme gradient boosting, neural networks with a single hidden layer, and naïve Bayes. For each algorithm, hyperparameters were optimised with 10 repeats of 10-fold cross-validation to maximise the area under the receiver operating characteristic curve (AUC) of the training model. The SMOTE subsampling method was utilised during training.

The performance of each prediction model was evaluated using the AUC in the internal/external testing/validation datasets. The optimal cut-off value for hip fracture risk classification was determined based on receiver operating characteristic analysis of the training dataset using Youden's index. The sensitivity, specificity, positive predictive value, negative predictive value, F1 statistic, accuracy, and error rate were evaluated. DeLong's test was used to compare AUCs. The net reclassification index was computed to assess whether predictions from one model significantly differed from those of another model.

Results

In total, the derivation cohort included 239 996 individuals. In the female cohort, 7.3% and 10.4% of individuals experienced hip fracture within 10- and 15-year follow-up periods, respectively. Fewer hip fracture cases were observed in the male cohort (3.7% and 5.4%, respectively). After exclusion of individuals lost to follow-up, the derivation cohort comprised 91 926 and 75 823 women and 69 125 and 52 345 men for 10- and 15-year prediction models, respectively (Tables 1a and 1b). Individuals in the external validation cohort were younger, had fewer hospitalisations prior to the index date, and had fewer hip fracture cases. Diagnosis and drug prescription records for some known risk factors of

TABLE 2b. Performance of hip fracture risk prediction models for women excluding those lost to follow-up

Performance	Stepwise selection by logistic regression		Gradient boosting machine		Random forest		Extreme gradient boosting		Neural networks with a single hidden layer		Naïve Bayes	
	10-year risk	15-year risk	10-year risk	15-year risk	10-year risk	15-year risk	10-year risk	15-year risk	10-year risk	15-year risk	10-year risk	15-year risk
Training cohort												
Area under the curve (95% confidence interval)	0.823 (0.818-0.827)	0.823 (0.819-0.828)	0.806 (0.801-0.811)	0.81 (0.805-0.814)	0.997 (0.997-0.998)	0.997 (0.996-0.997)	0.798 (0.792-0.803)	0.807 (0.802-0.811)	0.806 (0.801-0.811)	0.816 (0.811-0.82)	0.534 (0.529-0.54)	0.66 (0.654-0.667)
Testing cohort												
Area under the curve (95% confidence interval)	0.815 (0.805-0.825)	0.815 (0.806-0.824)	0.798 (0.788-0.809)	0.806 (0.797-0.815)	0.769 (0.758-0.78)	0.779 (0.769-0.789)	0.787 (0.776-0.7796)	0.804 (0.794-0.813)	0.784 (0.774-0.795)	0.806 (0.797-0.816)	0.634 (0.619-0.649)	0.658 (0.645-0.671)
Sensitivity	0.727	0.73	0.749	0.717	0.34	0.545	0.734	0.695	0.728	0.733	0.763	0.605
Specificity	0.748	0.74	0.707	0.742	0.923	0.847	0.717	0.761	0.709	0.739	0.312	0.71
Positive predictive value	0.248	0.376	0.227	0.374	0.337	0.434	0.229	0.384	0.223	0.377	0.113	0.31
Negative predictive value	0.96	0.927	0.961	0.924	0.924	0.896	0.959	0.921	0.958	0.928	0.92	0.893
F1	0.37	0.497	0.348	0.492	0.338	0.483	0.349	0.495	0.341	0.498	0.197	0.41
Accuracy	0.745	0.738	0.712	0.738	0.863	0.793	0.719	0.749	0.711	0.738	0.358	0.692
Error	0.255	0.262	0.288	0.262	0.137	0.207	0.281	0.251	0.289	0.262	0.642	0.308
External validation cohort												
Area under the curve (95% confidence interval)	0.841 (0.807-0.876)	0.845 (0.815-0.874)	0.845 (0.811-0.879)	0.843 (0.813-0.873)	0.773 (0.736-0.81)	0.783 (0.748-0.819)	0.837 (0.802-0.873)	0.84 (0.809-0.871)	0.806 (0.768-0.844)	0.838 (0.806-0.869)	0.595 (0.537-0.654)	0.64 (0.594-0.686)
Sensitivity	0.69	0.716	0.773	0.72	0.366	0.569	0.765	0.697	0.738	0.716	0.724	0.573
Specificity	0.814	0.818	0.762	0.813	0.908	0.838	0.77	0.823	0.777	0.803	0.238	0.745
Positive predictive value	0.221	0.349	0.199	0.344	0.233	0.324	0.203	0.348	0.202	0.331	0.068	0.234
Negative predictive value	0.972	0.955	0.978	0.955	0.949	0.935	0.977	0.952	0.975	0.954	0.918	0.928
F1	0.335	0.469	0.317	0.466	0.285	0.412	0.321	0.464	0.318	0.453	0.124	0.332
Accuracy	0.805	0.806	0.763	0.802	0.87	0.806	0.769	0.808	0.774	0.793	0.272	0.724
Error	0.195	0.194	0.237	0.198	0.13	0.194	0.231	0.192	0.226	0.207	0.728	0.276
Delong's test P value	Reference	Reference	0.469	0.618	<0.001	<0.001	0.556	0.351	<0.001	0.18	<0.001	<0.001
Integrated discrimination improvement (P value)	Reference	Reference	-0.428 (<0.001)	-0.517 (0.001)	-0.364 (<0.001)	-0.501 (0.001)	-0.445 (<0.001)	-0.533 (<0.001)	-0.455 (<0.001)	-0.567 (<0.001)	-0.089 (0.005)	-0.225 (<0.001)

fracture were less prevalent.

In the female prediction model, 239 potential predictors were used. The LR approach had the highest AUC in the external validation cohort for the 10-year (0.769) and 15-year (0.724) prediction models (Table 2a). After exclusion of individuals lost to follow-up, 233 potential predictors were used. The gradient boosting machine model had the highest AUC in external validation (0.841) for predicting 10-year risk, but DeLong's test showed no significant difference relative to LR (Table 2b). Both the LR approach and gradient boosting machine model achieved moderate sensitivity and specificity (>0.70). For predicting 15-year risk, the LR approach provided the best AUC (0.845) in the external validation cohort, attaining high specificity (0.818) and moderate sensitivity (0.716).

DeLong's test showed that the LR approach had a significantly higher AUC compared with the random forest, neural networks, and naïve Bayes models. Using the LR approach as reference, all other prediction models displayed significant and negative integrated discrimination improvement, indicating that the LR approach had better discrimination performance in predicting hip fracture risk. The LR approach identified 20 risk factors for women: age, number of hospitalisations, and diagnosis/drug prescription variables for accidental falls, heart disease, diabetes, Parkinson's disease, chronic kidney disease, psychoses, chronic obstructive pulmonary disease, depression, epilepsy, nutritional deficiencies, and history of fracture.

In the male prediction model, 238 potential predictors were used. For predicting 10-year risk, the LR approach had the highest AUC (0.805) in the external validation cohort; its sensitivity and specificity were moderate (>0.7). The LR approach also yielded the best AUC (0.723) in external validation for predicting the 15-year risk, although its sensitivity was lower than that of the 10-year

TABLE 3a. Performance of hip fracture risk prediction models for men.

Performance	Stepwise selection by logistic regression		Gradient boosting machine		Random forest		Extreme gradient boosting		Neural networks with a single hidden layer		Naïve Bayes	
	10-year risk	15-year risk	10-year risk	15-year risk	10-year risk	15-year risk	10-year risk	15-year risk	10-year risk	15-year risk	10-year risk	15-year risk
Training cohort												
Area under the curve (95% confidence interval)	0.726 (0.717-0.734)	0.668 (0.661-0.676)	0.665 (0.656-0.675)	0.614 (0.605-0.622)	0.996 (0.995-0.997)	0.995 (0.994-0.995)	0.645 (0.635-0.655)	0.595 (0.587-0.603)	0.689 (0.68-0.697)	0.652 (0.644-0.66)	0.507 (0.497-0.518)	0.504 (0.495-0.513)
Testing cohort												
Area under the curve (95% confidence interval)	0.703 (0.687-0.72)	0.673 (0.658-0.688)	0.664 (0.645-0.683)	0.628 (0.612-0.644)	0.625 (0.608-0.643)	0.575 (0.56-0.591)	0.644 (0.625-0.644)	0.607 (0.591-0.624)	0.678 (0.661-0.695)	0.651 (0.635-0.667)	0.50 (0.5-0.5)	0.497 (0.48-0.514)
Sensitivity	0.705	0.642	0.658	0.733	0.05	0.097	0.556	0.627	0.81	0. 662	0.642	0.627
Specificity	0.614	0.613	0.611	0.482	0.963	0.919	0.68	0.55	0.476	0.578	0.352	0.363
Positive predictive value	0.066	0.0874	0.062	0.075	0.049	0.064	0.063	0.07314	0.057	0.082	0.037	0.053
Negative predictive value	0.982	0.968	0.979	0.969	0.963	0.947	0.975	0.963	0.985	0.968	0.962	0.945
F1	0.121	0.153	0.113	0.136	0.049	0.077	0.114	0.132	0.106	0.147	0.07	0.098
Accuracy	0.617	0.614	0.613	0.496	0.929	0.875	0.676	0.554	0.488	0.583	0.363	0.377
Error	0.383	0.386	0.387	0.504	0.071	0.125	0.324	0.446	0.512	0.418	0.637	0.622
External validation cohort												
Area under the curve (95% confidence interval)	0.805 (0.734-0.876)	0.723 (0.655-0.791)	0.71 (0.602-0.818)	0.684 (0.607-0.761)	0.655 (0.594-0.717)	0.62 (0.548-0.692)	0.692 (0.583-0.802)	0.672 (0.59-0.754)	0.701 (0.62-0.783)	0.707 (0.633-0.78)	0.5 (0.5-0.5)	0.55 (0.461-0.639)
Sensitivity	0.75	0.69	0.722	0.741	0.167	0.362	0.694	0.741	0.806	0.672	0.667	0.724
Specificity	0.718	0.716	0.6547	0.494	0.895	0.783	0.712	0.514	0.514	0.606	0.239	0.25
Positive predictive value	0.072	0.104	0.057	0.065	0.044	0.074	0.065	0.0676	0.046	0.075	0.025	0.044
Negative predictive value	0.99	0.98	0.988	0.976	0.974	0.963	0.988	0.977	0.989	0.975	0.961	0.95
F1	0.131	0.18	0.106	0.12	0.07	0.122	0.12	0.124	0.087	0.135	0.048	0.083
Accuracy	0.719	0.715	0.656	0.505	0.875	0.764	0.712	0.524	0.522	0.609	0.251	0.272
Error	0.281	0.285	0.344	0.495	0.125	0.236	0.288	0.476	0.478	0.391	0.749	0.728
Delong's test P value	Reference	Reference	0.04	0.275	<0.001	0.023	0.024	0.209	0.013	0.523	<0.001	0.002
Integrated discrimination improvement (P value)	Reference	Reference	-0.135 (<0.001)	-0.083 (<0.001)	-0.082 (<0.001)	-0.109 (<0.001)	-0.136 (<0.001)	-0.088 (<0.001)	-0.171 (<0.001)	-0.138 (<0.001)	-0.036 (<0.001)	-0.022 (<0.001)

prediction model (Table 3a). After exclusion of individuals lost to follow-up, 233 potential predictors were included in model development. The LR approach displayed the best AUCs in external validation: 0.898 for 10 years and 0.843 for 15 years. Both sensitivity and specificity were >0.8 in predicting the 10-year risk, but the sensitivity for the 15-year prediction model dropped below 0.7 (Table 3b).

DeLong's test showed that the LR approach had a significantly higher AUC compared with the random forest and naïve Bayes models. The LR approach had better discrimination performance in predicting hip fracture risk; all other prediction models showed significant and negative integrated discrimination improvement with reference to the LR approach. The LR approach identified 20 risk factors for men: age, number of hospitalisations, and diagnosis/drug prescription variables for heart disease, diabetes, Parkinson's disease, chronic kidney disease, psychoses, chronic obstructive pulmonary disease, depression, epilepsy, nutritional deficiencies, and history of fracture.

Discussion

In a German osteoporotic hip fracture prediction model involving 288 086 individuals, age, sex, history of fracture, and medication were identified as predictors. The model had an AUC of 0.65 to 0.7.[5] The use of a pre-defined set of risk factors to train the prediction model may exclude strong risk factors not collected in past studies. Therefore, our study included all diagnosis and drug prescription records as potential predictors. After exclusion of individuals lost to follow-up, our best-performing prediction models attained AUCs >0.8 in external validation, indicating clinical utility. Notably, the external validation cohort comprised community-dwelling individuals, demonstrating the high generalisability

TABLE 3b. Performance of hip fracture risk prediction models for men excluding those lost to follow-up

Performance	Stepwise selection by logistic regression		Gradient boosting machine		Random forest		Extreme gradient boosting		Neural networks with a single hidden layer		Naïve Bayes	
	10-year risk	15-year risk	10-year risk	15-year risk	10-year risk	15-year risk	10-year risk	15-year risk	10-year risk	15-year risk	10-year risk	15-year risk
Training cohort												
Area under the curve (95% confidence interval)	0.826 (0.819-0.834)	0.819 (0.812-0.825)	0.796 (0.788-0.804)	0.796 (0.79-0.803)	0.997 (0.996-0.998)	0.996 (0.995-0.997)	0.785 (0.777-0.794)	0.789 (0.782-0.796)	0.818 (0.81-0.825)	0.811 (0.804-0.817)	0.519 (0.512-0.526)	0.647 (0.638-0.657)
Testing cohort												
Area under the curve (95% confidence interval)	0.817 (0.801-0.834)	0.815 (0.801-0.829)	0.796 (0.779-0.813)	0.797 (0.783-0.812)	0.763 (0.746-0.78)	0.767 (0.752-0.781)	0.785 (0.767-0.802)	0.794 (0.779-0.808)	0.798 (0.782-0.815)	0.797 (0.783-0.812)	0.637 (0.613-0.66)	0.65 (0.631-0.669)
Sensitivity	0.743	0.748	0.719	0.708	0.321	0.484	0.707	0.692	0.73	0.722	0.801	0.564
Specificity	0.75	0.741	0.739	0.763	0.937	0.886	0.75	0.778	0.737	0.752	0.232	0.755
Positive predictive value	0.159	0.272	0.149	0.278	0.244	0.353	0.152	0.287	0.15	0.274	0.062	0.229
Negative predictive value	0.979	0.958	0.976	0.953	0.956	0.93	0.976	0.951	0.977	0.954	0.948	0.931
F1	0.262	0.399	0.246	0.399	0.277	0.408	0.251	0.406	0.249	0.397	0.115	0.326
Accuracy	0.75	0.742	0.738	0.757	0.9	0.84	0.748	0.768	0.737	0.749	0.266	0.733
Error	0.25	0.258	0.262	0.243	0.1	0.16	0.252	0.232	0.263	0.251	0.734	0.267
External validation cohort												
Area under the curve (95% confidence interval)	0.898 (0.857-0.939)	0.843 (0.788-0.898)	0.854 (0.787-0.921)	0.84 (0.7854-0.896)	0.758 (0.692-0.825)	0.775 (0.711-0.84)	0.809 (0.72-0.897)	0.818 (0.759-0.878)	0.865 (0.796-0.935)	0.832 (0.77-0.893)	0.573 (0.444-0.702)	0.678 (0.596-0.759)
Sensitivity	0.806	0.707	0.722	0.707	0.306	0.483	0.778	0.672	0.778	0.707	0.667	0.534
Specificity	0.816	0.802	0.771	0.786	0.882	0.865	0.776	0.8	0.798	0.781	0.199	0.816
Positive predictive value	0.139	0.21	0.104	0.198	0.087	0.211	0.114	0.201	0.125	0.194	0.03	0.178
Negative predictive value	0.991	0.973	0.987	0.973	0.972	0.957	0.99	0.97	0.99	0.973	0.941	0.959
F1	0.238	0.324	0.182	0.309	0.136	0.293	0.199	0.31	0.215	0.305	0.057	0.267
Accuracy	0.816	0.795	0.769	0.781	0.861	0.838	0.776	0.791	0.798	0.776	0.215	0.796
Error	0.184	0.205	0.231	0.219	0.139	0.162	0.224	0.209	0.202	0.224	0.785	0.204
Delong's test P value	Reference	Reference	0.102	0.806	<0.001	0.027	0.023	0.15	0.211	0.5	<0.001	<0.001
Integrated discrimination improvement (P value)	Reference	Reference	-0.427 (<0.001)	-0.435 (<0.001)	-0.339 (<0.001)	-0.426 (<0.001)	-0.434 (<0.001)	-0.445 (<0.001)	-0.526 (<0.001)	-0.502 (<0.001)	-0.117 (0.025)	-0.271 (<0.001)

of our prediction models. Additionally, the prediction model did not use BMD as a predictor, owing to the limited availability of dual-energy X-ray absorptiometry in Hong Kong.

Using a data-driven approach that does not rely on known associations between predictor variables and fracture, some novel predictors were identified. One example was the diagnosis or drug prescription for anaemias and other blood disorders, which were associated with higher odds of hip fracture. This finding is consistent with the results of our Mendelian randomisation study, which demonstrated a positive causal association of genetically determined red blood cell traits with BMD. Individuals with haematological diseases (eg, anaemia) may have higher lifelong risks of osteoporosis and fracture. Other novel predictors, such as the use of emollient and barrier preparations, laxatives, and vitamins/minerals, were indicators of ageing or frailty; they were usually prescribed for dry skin, constipation, and poor appetite. The underlying mechanisms of how these novel predictors affect fracture risk warrant future investigation.

Model performance may be further improved by including medical records from >1 year prior to the index date, and by incorporating medical information such as laboratory test results and surgical procedures. With additional validation in independent cohorts from the Hong Kong population, these prediction models may serve as routine screening tools in future public healthcare settings.

Conclusions

We developed ethnicity- and sex-specific hip fracture prediction models for the Hong Kong population using machine learning algorithms and electronic medical records. The prediction models demonstrated good performance, achieving AUCs

>0.8. The prediction models may be clinically useful and generalisable to the public.

Funding

This study was supported by the Health and Medical Research Fund, Health Bureau, Hong Kong SAR Government (#17181381). The full report is available from the Health and Medical Research Fund website (https://rfs2.healthbureau.gov.hk).

Disclosure

The results of this research have been previously published in:

1. Li GH, Cheung CL, Tan KC, et al. Development and validation of sex-specific hip fracture prediction models using electronic health records: a retrospective, population-based cohort study. EClinicalMedicine 2023;58:101876.

References

1. 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;4:16-21.
2 Cheung EY, Bow CH, Cheung CL, et al. Discriminative value of FRAX for fracture prediction in a cohort of Chinese postmenopausal women. Osteoporos Int 2012;23:871-8.
3. International Osteoporosis Foundation. Asia-Pacific Regional Audit. Accessed 6 March 2025. Available from: https://www.osteoporosis.foundation/sites/iofbonehealth/files/2019-06/2013_Asia_Pacific_Audit_English.pdf.
4. Cheung CL, Tan KCB, Kung AWC. Cohort profile: the Hong Kong Osteoporosis Study and the follow-up study. Int J Epidemiol 2018;47:397-8f.
5. 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:e0232969.