



Development and validation of a novel screening tool for osteoporosis in older US adults: The NHANES cross-sectional study

Yuchen Tang^{1,2,3} · Zhongcheng Liu^{1,2,3} · Shenghong Wang^{1,2,3} · Qiong Yi^{1,2,3} · Yayi Xia^{1,2,3} · Bin Geng^{1,2,3}

Received: 14 September 2021 / Accepted: 29 January 2022 / Published online: 5 February 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Purpose Osteoporosis (OP) is a common disease among adults aged >50 years. At present, the main approach to screen or to diagnosis OP is mainly via bone mineral density (BMD) testing, which might not be optimal for OP screening. This study aimed to develop and validate a convenient and effective prediction model for screening OP based on the demographic information, medical history, and lifestyle habits in the elderly in the United States.

Methods All data were collected from the National Health and Nutrition Survey database. Participants aged ≥50 years with complete BMD data were included in this study. Twelve candidate predictors were initially selected to develop the prediction model. Final predictors screening and model development were based on multivariate logistic regression. Model discrimination (C statistic) and calibration (Brier scores) were calculated to evaluate the performance of the model. Internal validation was performed using the bootstrap resampling technique, and external validation was based on the validation cohort.

Results The screening tool was developed with individual patient data from 1941 patients and validated with data from 1947 patients after the development of the model. Seven predictors (patient age, sex, race, body mass index, physical activity, sleep duration, and history of fracture) were included in the final prediction model, and the final model had a C statistic of 0.849 [95% confidence interval (CI): 0.820–0.878] and Brier scores of 0.062 [95% CI: 0.054–0.070] on the development cohort. For the validation of the developed model, the results showed a C statistic >0.800 and Brier scores <0.070, irrespective of internal validation or external validation.

Conclusions A novel screening tool for OP in the elderly, which has excellent discrimination and useful calibration, has been developed and externally validated. Considering its simplicity, generalizability, and accuracy, this tool has the potential to become a practical mean for the elderly to screen OP.

Keywords Osteoporosis · Prediction model · The elderly · National Health and Nutrition Survey · External validation

These authors contributed equally: Yuchen Tang, Zhongcheng Liu

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1007/s12020-022-03001-2>.

✉ Bin Geng
cxxx@foxmail.com

¹ Department of Orthopaedics, Lanzhou University Second Hospital, Lanzhou, Gansu, China

² Orthopaedics Key Laboratory of Gansu Province, Lanzhou, Gansu, China

³ Orthopaedic Clinical Research Center of Gansu Province, Lanzhou, Gansu, China

Introduction

Osteoporosis (OP), characterized by reduced bone mineral density (BMD) and deterioration in bone microarchitecture, is a chronic disease with a high incidence worldwide [1, 2]. According to the International Osteoporosis Foundation (IOF), approximately one-fifth of men and one-third of women, who are aged >50 years, are at risk of OP and osteoporotic fractures [2]. Management and monitoring of disease always play essential roles in the health of the elderly, such as glucose monitoring for diabetes and blood pressure monitoring for hypertension [3, 4], which are both crucial components for the screening of disease and easily implemented in daily life situations. However, the current diagnosis of OP is based on BMD testing [1], which might not be optimal for OP screening. Other potential methods for OP diagnosis, such as bone turnover markers need to be performed in hospitals or

other health care organizations [1]. Therefore, there is a great need to develop a novel method with high accuracy and full maneuverability to manage OP in the elderly.

The clinical prediction model has received increasing attention in recent years, and accumulating evidence has demonstrated that the prediction model has potential clinical utility in disease diagnosis or screening. For example, Jaja et al. developed and validated a prediction model for aneurysmal subarachnoid hemorrhage with the area under the receiver operator characteristic (ROC) curve (AUC) of 0.76 for full model [5]. Similarly, Vistisen et al. developed and validated a novel prediction model for the first cardiovascular disease event in patients with type-1 diabetes (AUC of 0.826 for the derivation data and 0.803 for the validation data) [6]. Regarding OP, most assessment tools or prediction models were used to assess or predict the risk of osteoporotic fractures like FRAX, FEM, or QFracture [7–9]. For example, Kanis et al. introduced a fracture risk assessment tool (FRAX), the most commonly used tool to assess fracture risk until today [8]. Rubin et al. reported that the FEM tool for predicting major or hip osteoporotic fractures showed high accuracy for individuals aged ≥ 45 years (For major osteoporotic fractures: AUC of 0.750 and 0.752 for women and men; For hip fractures: AUC of 0.874 and 0.851 for women and men) [9]. Hippisley-Cox et al. introduced a new tool (QFracture) to estimate the risk of osteoporotic fracture or hip fracture over 10 years, which also showed good performance for discrimination and calibration (For major osteoporotic fractures: AUC of 0.79 and 0.69 for women and men; For hip fractures: AUC of 0.89 and 0.86 for women and men) [7]. For the screening of osteoporosis, Wang et al. developed a prediction model for OP based on the neural network [10], which showed an AUC of 0.901 in the training set and 0.762 in the test set. Ho-Pham et al. also developed a prediction model for OP based on logistic regression [11]. Although these two models showed good performance, these studies lacked external validation. Koh et al. developed a simple tool (OSTA) to predict the risk of OP with 91% sensitivity and 45% specificity [12]. Interestingly, Ma et al. also developed a new predictive screening tool (BFH-OST) for OP screening that had a sensitivity of 73.6% and specificity of 72.7% [13]. Moreover, Ma et al. compared the performance of the two aforementioned methods and found that BFH-OST performed better than OSTA [13]. However, those two methods mentioned above had limited predictive value for the risk of OP, which requires further improvement. Furthermore, both methods were established based on the specific populations. Although OSTA or BFH-OST appears to be a simple, rapid, and convenient method, some potential factors that have been demonstrated to affect bone metabolism, such as sleep duration [14], physical activity [15], or glucocorticoids [16], might be neglected. Therefore,

considering the limitations of these existing models and the continuous renewal of relevant knowledge in OP, it remains necessary to develop a novel screening tool with high accuracy for OP management.

Taking into account the actual demands for OP management and the limitations of the existing literature mentioned above, this study aimed to develop a novel screening tool for OP in the elderly based on demographic information, medical history, and lifestyle habits. Moreover, the performance of the prediction model was evaluated and validated for generalization and application.

Materials and methods

Study population and data collection

All participant information was collected independently by two investigators from the National Health and Nutrition Survey (NHANES) database, which is affiliated with the US Centers for Disease Control and Prevention [17]. The NHANES is a national program that aims to assess the health and nutritional status of community-living American residents both adults and children [17]. The NHANES database collects and stores information related to interviews and examinations and is updated biennially [17]. The ethics review board of the National Center for Health Statistics approved the study. Written informed consent was obtained from each participant before their inclusion on the NHANES database. Detailed information on the ethics application and written informed consent are provided on the NHANES website [18].

Data from the NHANES during 2013–2014 were adopted as the development cohort [19]. The NHANES data during 2017–2018 were adopted as the validation cohort [20]. Two investigators independently extracted information from the database on July 1, 2021. All pertinent data collected were as follows:

- i. Participant demographic data: age, sex, and race.
- ii. Examination data: Total femur BMD (determined by dual-energy X-ray absorptiometry), femoral neck BMD (determined by dual-energy X-ray absorptiometry), trochanter BMD (determined by dual-energy X-ray absorptiometry), and intertrochanter BMD (determined by dual-energy X-ray absorptiometry).
- iii. Questionnaire data: height (current self-reported height); weight (current self-reported weight); smoking status (smoked at least 100 cigarettes or not in life); alcohol consumption (had at least 12 alcohol drinks in the past year); diabetes (previous diagnosis by a doctor); physical activity (including vigorous work-related activity, moderate work-related activity,

walking or bicycling for transportation, vigorous leisure-time physical activity, and moderate leisure-time physical activity); sleep duration (how much sleep on average); prednisone or cortisone use (prednisone or cortisone daily); family history of OP (either biological parent diagnosed by a health professional as having OP or brittle bones); and previous fractures (previous diagnosis of broken or fractured hip, wrist, spine, or any other bone).

In addition, the baseline characteristics of outcome measures and candidate predictors in the development cohort and validation cohort were listed as mean for continuous variables and as proportion for categorical variables. Additional details about the data extraction are listed in Supplementary Table S1.

Inclusion criteria

Participants who were aged >50 years with complete BMD data (all sites of the femur) were included in the development and validation of the prediction model. The NHANES 2013–2014 comprised 10,175 participants, while the NHANES 2017–2018 comprised 9254 participants. All subjects aged <50 years were excluded (NHANES 2013–2014: $n = 7395$; NHANES 2017–2018: $n = 6185$). Next, subjects without complete BMD information (any site of the femur) were excluded (NHANES 2013–2014: $n = 550$; NHANES 2017–2018: $n = 783$). Subsequently, 2230 participants (NHANES 2013–2014) were enrolled in the development cohort, while 2286 participants (NHANES 2017–2018) were enrolled in the validation cohort. The flowchart of participant selection is shown in Fig. 1.

Candidate predictors selection

The selection of candidate predictors was mainly based on clinical guidelines, related studies on the risk factors of OP

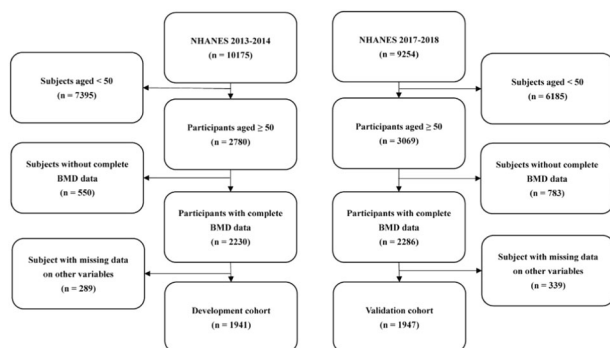


Fig. 1 Participant selection for development cohort and validation cohort. NHANES National Health and Nutrition Survey, BMD bone mineral density

[searched PubMed, Embase, and Cochrane Library databases according to preferred reporting items for systematic reviews and meta-analyses guidelines] [21], IOF Osteoporosis Risk Check [2], and expert opinion. The main principles of selection were: (i) data that could be collected from the NHANES database; (ii) data that excluded any laboratory indicators; and (iii) data that were easily accessible in real applications.

The selection of candidate predictors was based on the rules of the number of events per variable in the logistic regression analysis [22]. All information on candidate predictors was assessed by the researchers of the NHANES program, who were not collaborators in this study. Considering the generalizability and applicability of the prediction model, all candidate predictors selected were collected via face-to-face or telephone interviews. The details of the assessment of any candidate predictor can be found in the NHANES database [17].

Twelve candidate predictors selected included: age (continuity variable; years); sex (categorical variable; male or female); race (categorical variable; Mexican American, other Hispanic, non-Hispanic white, non-Hispanic black, or other races); body mass index (BMI; categorical variable; normal: <25 kg/m², overweight: ≥25 and <30 kg/m², or obesity: ≥30 kg/m²); smoking status (categorical variable; yes or no); alcohol consumption (categorical variable; yes or no); diabetes (categorical variable; yes, no or borderline); physical activity (categorical variable; no: 0 metabolic equivalent of task [MET]-mins/week, low: 1–<600 MET-mins/week, medium: 600–<1200 MET-mins/week, and high: ≥1200 MET-mins/week according to the World Health Organization (WHO) PA guidelines and the method employed in a previous study [23, 24]. Suggested MET score for vigorous work-related activity or vigorous leisure-time physical activity was 8.0, while suggested MET score for moderate work-related activity, walking or bicycling for transportation, and moderate leisure-time physical activity was 4.0 according to the information provided on the NHANES website) [25]; sleep duration (categorical variable; short: <7 h; normal: between 7 and 9 h; long: >9 h according to the study by Hirshkowitz et al.) [26]; prednisone or cortisone use (categorical variable; yes or no); family history of OP (categorical variable; yes or no); and previous fractures (categorical variable; yes or no).

Outcome measure

BMD testing was performed using dual-energy X-ray absorptiometry by technologists who were certified radiology technologists and were not collaborators in this study. This study only collected BMD data of the total femur, femur neck, trochanter, and intertrochanter. Detailed information on the protocol and procedure of BMD testing is provided on the NHANES website [27].

In the present study, we collected the BMD data from the NHANES database, and OP was defined as a BMD value in any femoral region (total femur, femur neck, trochanter, or intertrochanter), measuring < -2.5 standard deviations from the reference value. Moreover, Looker et al. recommended using the mean femoral BMD for non-Hispanic white men and women aged 20–29 years from the NHANES III database as the BMD reference value according to a previous study [28]. The specific reference values are listed in Supplementary Table S2. Two researchers independently assessed the outcome (OP or non-OP) for all participants included in the final analysis by the information on BMD as a part of this project, and any discrepancies in outcome assessment were resolved by discussion with a third researcher.

Handling of missing data

Except for any missing information, any questions wherein participants refused to answer or answered “I do not know” (if the choice existed) were also considered to be missing data. All procedures for the development or validation of the prediction model were based on a cohort with complete data (including outcome measures and candidate predictors). Participants with missing data for any candidate predictor were excluded from the final analysis. The number of participants enrolled in the final analysis and the subjects with missing data in the development and validation cohorts are shown in Supplementary Fig. S1. Finally, 628 subjects (289 from the NHANES 2013–2014; 339 from the NHANES 2017–2018) with missing data on other variables were excluded, and 3888 participants (1941 in the development cohort and 1947 in the validation cohort) were included in the final analysis.

Model development

The final predictor selection for model development was performed using bidirectional elimination by running stepwise regression on all model combinations using all candidate predictors with *P*-value criteria for retention set at 0.20 according to some previous study [29, 30]. The Student *t*-test (for continuity variable) or chi-square test (for categorical variable) was performed one by one for the selected variables. When the original variable was no longer significant owing to the introduction of later variables, the original variable was deleted. This procedure was repeated until neither a significant variable was selected into the regression equation, nor the insignificant variable was removed from the regression equation. The final model to predict whether OP occurred was derived based on the final selected variables' coefficients based on the logistic regression model. Moreover, the optimal cutoff value was

determined by the maximum Youden index (Youden index = sensitivity + specificity – 1). In addition, the final prediction model was displayed by an interactive web application, which was based on the R package Shiny [31].

Model performance

The model performance of the developed prediction model was evaluated in two aspects:

- (i) Model discrimination: the AUC (the AUC is equal to the value of the C statistic) was calculated to assess the ability of the prediction model to differentiate between participants who did or did not have OP;
- (ii) Model calibration: Calibration plots and Brier scores were used to indicate the ability of the models to produce unbiased estimates of the probability of the outcome.

Model validation

Model validation mainly included two procedures as follows:

- (i) Internal validation: Internal validation was performed using the bootstrap resampling technique, in which regression models were fitted in 500 bootstrap replicates, drawn with replacement from the development sample. The model was refitted in each bootstrap replicate (repeated the predictors selection and model fit) and tested on the original sample to estimate optimism in model performance.
- (ii) External validation: External validation was performed with the validation datasets.

Statistical software

All analyses were performed with R software, V.4.0.5 [R: a language and statistical computing environment (program). Vienna, Austria; R Foundation for Statistical Computing, 2016] and EmpowerStats (<http://www.empowerstats.com>).

Results

Baseline characteristics

Overall, 3888 participants (1941 in the development cohort and 1947 in the validation cohort) were included in the final analysis. There were 151 participants with OP in the development cohort and 165 subjects with OP in the validation

cohort. Compared to subjects without osteoporosis (non-OPs), participants with OP were more likely to be older, female, non-Hispanic Whites, other races, or normal-weight participants. Moreover, the percentage of participants who smoked <100 cigarettes in their whole life had <12 times drinking during the past year, without diabetes, without moderate-to-vigorous physical activity, and long sleep duration in the OPs were higher than those in the non-OP group. In addition, OPs were more likely to use prednisone or cortisone, had a family history of OP, and had a history of fracture compared with non-OPs. Detailed results are presented in Table 1.

Model development

For predictor selection, of the 12 candidate predictors, eight predictors (age, sex, race, BMI, alcohol consumption status, physical activity, sleep duration, and previous fracture) were selected to develop the prediction model using bidirectional elimination by running stepwise regression (Supplementary Table S3). The coefficient for subjects who had at least 12 times drinking during the past year was -0.32882 , which might suggest that drinking more alcohol was a protective factor for OP, and this finding contradicts previous conclusions that alcohol consumption was positively associated with the risk of OP [32, 33]. Moreover, we analyzed the age and sex ratio for different alcohol consumption statuses, and participants who consumed less time tended to be older and more often female (Supplementary Table S4), which was a plausible explanation for this result. In addition, the coefficient of this model suggested that individuals with low MVPA had a lower risk of osteoporosis compared with those with medium or high MVPA, which was not in line with findings from the previous studies and might cause misleading results [34, 35]. We next ran the analysis for predictor selection with 11 candidate predictors (except for alcohol consumption status). Moreover, we modified the grouping criterion by dividing all participants into three groups (no MVPA, low or medium MVPA, and high MVPA). Finally, seven predictors (age, sex, race, BMI, physical activity, sleep duration, and previous fracture) were finally selected to develop the prediction model (Table 2). For model performance, the final model showed an AUC of 0.849 (95% CI: 0.820–0.878) and Brier scores of 0.062 (95% CI: 0.054–0.070). The optimal cutoff of the predicted probability was 0.072 (sensitivity, 0.841; specificity, 0.729). The local maxima of the ROC curve are listed in (Supplementary Table S5). Moreover, the ROC curve, calibration plot, and other results of model performance are shown in Fig. 2a, c, and Table 3, respectively.

Internal validation

Internal validation was performed using bootstrap resampling. After 500 repetitions, the test performance showed an

AUC of 0.838 (95% CI: 0.837–0.838) and Brier scores of 0.063 (95% CI: 0.063–0.063). Moreover, the average optimism for AUC and Brier scores was $+0.017$ and -0.002 , respectively. Finally, the AUC and Brier scores after optimism adjustment were 0.833 (95% CI: 0.831–0.834) and 0.064 (95% CI: 0.064–0.064), respectively. The results are presented in Table 3.

External validation

The results of external validation showed an AUC of 0.807 (95% CI: 0.775–0.839) and Brier scores of 0.068 (95% CI: 0.059–0.077). The ROC curve, calibration plot, and other results of model performance are shown in Fig. 2b, d, and Table 3, respectively.

Model presentation

The final model constructed using the development cohort was presented by an interactive web application, which was based on the R package Shiny [31]. The URL of this web application is <https://yuchentanglzu1994.shinyapps.io/dynnomapp/>, where other researchers can replicate the analysis and validate the performance of this model. A screenshot of the model summary is shown in Fig. 3. As an example case, given a woman aged 72 years, non-Hispanic White, with a normal BMI, no MVPA, short sleep duration, and a history of hip fracture, the result showed that this woman had a 0.630 probability (95% CI: 0.506–0.739) for OP (Fig. 4).

Discussion

This study has developed a novel risk prediction model for OP in older United States (US) adults, which showed excellent discrimination and good calibration. Moreover, this study also validated the performance of this model, with results showing a C statistic >0.800 and Brier scores <0.070 , irrespective of internal validation or external validation. In addition, the risk prediction model was presented by an interactive web application, which was designed specifically to be easy to use or to replicate this analysis.

Eight of the 12 predictors were finally included in model development using bidirectional elimination by running stepwise regression for predictor selection. In comparison, the other four predictors (smoking status, diabetes, prednisone or cortisone use, and family history of osteoporosis), which were considered risk factors for OP in previous studies [1, 32, 33], were excluded. First, we considered that the information collected might not completely reflect the actual condition. For example, the use of prednisone or cortisone has been considered a risk factor for

Table 1 Characteristics of development cohort and validation cohort

Characteristic	Mean or proportion			
	Development cohort		Validation cohort	
	Non-osteoporosis (<i>N</i> = 1790)	Osteoporosis (<i>N</i> = 151)	Non-osteoporosis (<i>N</i> = 1782)	Osteoporosis (<i>N</i> = 165)
Age, mean (SD)	63.5 (9.0)	70.2 (8.7)	63.9 (8.7)	70.9 (9.1)
Gender, <i>n</i> (%)				
Male	933 (52.1)	46 (30.5)	995 (55.8)	50 (30.3)
Female	857 (47.9)	105 (69.5)	787 (44.2)	115 (69.7)
Race, <i>n</i> (%)				
Mexican American	219 (12.2)	10 (6.6)	196 (11.0)	7 (4.2)
Other hispanic	166 (9.3)	9 (6.0)	186 (10.4)	13 (7.9)
Non-hispanic white	813 (45.4)	100 (66.2)	650 (36.5)	103 (62.4)
Non-hispanic black	407 (22.7)	7 (4.6)	454 (25.5)	13 (7.9)
Other Races	185 (10.3)	25 (16.6)	296 (16.6)	29 (17.6)
BMI, <i>n</i> (%)				
Normal	503 (28.1)	89 (58.9)	469 (26.3)	90 (54.5)
Overweight	710 (39.7)	44 (29.1)	711 (39.9)	48 (29.1)
Obesity	577 (32.2)	18 (11.9)	602 (33.8)	27 (16.4)
Smoking status, <i>n</i> (%)				
No	892 (49.8)	86 (57.0)	939 (52.7)	93 (56.4)
Yes	898 (50.2)	65 (43.0)	843 (47.3)	72 (43.6)
Alcohol consumption status, <i>n</i> (%)				
No	506 (28.3)	64 (42.4)	1054 (59.1)	119 (72.1)
Yes	1284 (71.7)	87 (57.6)	728 (40.9)	46 (27.9)
Diabetes, <i>n</i> (%)				
No	1369 (76.5)	119 (78.8)	1294 (72.6)	134 (81.2)
Yes	337 (18.8)	28 (18.5)	409 (23.0)	29 (17.6)
Borderline	84 (4.7)	4 (2.6)	79 (4.4)	2 (1.2)
Physical activity, <i>n</i> (%)				
No MVPA	487 (27.2)	76 (50.3)	512 (28.7)	71 (43.0)
Low MVPA	325 (18.2)	16 (10.6)	243 (13.6)	24 (14.5)
Medium MVPA	210 (11.7)	17 (11.3)	201 (11.3)	15 (9.1)
High MVPA	768 (42.9)	42 (27.8)	826 (46.4)	55 (33.3)
Sleep duration, <i>n</i> (%)				
Normal	1107 (61.8)	82 (54.3)	1128 (63.3)	96 (58.2)
Short	623 (34.8)	59 (39.1)	462 (25.9)	40 (24.2)
Long	60 (3.4)	10 (6.6)	192 (10.8)	29 (17.6)
Prednisone or cortisone use, <i>n</i> (%)				
No	1690 (94.4)	139 (92.1)	1659 (93.1)	153 (92.7)
Yes	100 (5.6)	12 (7.9)	123 (6.9)	12 (7.3)
Family history of osteoporosis, <i>n</i> (%)				
No	1553 (86.8)	117 (77.5)	1548 (86.9)	126 (76.4)
Yes	237 (13.2)	34 (22.5)	234 (13.1)	39 (23.6)
Previous fracture, <i>n</i> (%)				
No	1082 (60.4)	63 (41.7)	1213 (68.1)	88 (53.3)
Yes	708 (39.6)	88 (58.3)	569 (31.9)	77 (46.7)

Abbreviations: *BMI* Body mass index, *MVPA* Moderate-to-vigorous physical activity, *SD* Standard deviation, *n* numbers of subjects, % Percentage

Table 2 Final multivariable analysis for osteoporosis risk in development cohort

Index	Estimate	Std. Error	z value	Pr(> z)
Intercept	-7.40845	0.874146	-8.475	<2e-16
Age	0.074304	0.01096	6.779	1.21E-11
Gender (male)	-	-	-	-
Gender (female)	0.714135	0.198511	3.597	0.000321
Race (Mexican American)	-	-	-	-
Race (Other hispanic)	0.008722	0.499744	0.017	0.986075
Race (Non-hispanic white)	0.536415	0.367315	1.46	0.144189
Race (Non-hispanic black)	-1.234304	0.525339	-2.35	0.018797
Race (Other races)	0.849713	0.419293	2.027	0.04271
BMI (Normal)	-	-	-	-
BMI (Overweight)	-1.054545	0.211012	-4.998	5.81E-07
BMI (Obesity)	-1.649777	0.287933	-5.73	1.01E-08
Physical activity (No MVPA)	-	-	-	-
Physical activity (Low or Medium MVPA)	-0.985065	0.236807	-4.16	3.19E-05
Physical activity (High MVPA)	-1.055276	0.224114	-4.709	2.49E-06
Sleep duration (Normal)	-	-	-	-
Sleep duration (Short)	0.621129	0.201561	3.082	0.002059
Sleep duration (Long)	0.380614	0.402901	0.945	0.34482
Previous fracture (No)	-	-	-	-
Previous fracture (Yes)	0.719294	0.191825	3.75	0.000177

Abbreviations: *Std* Standard deviation, *BMI* Body mass index, *MVPA* Moderate-to-vigorous physical activity

osteoporosis, but this variable was not included in the final model after predictor selection. We considered that the information on the use of prednisone or cortisone was based on the questionnaire, in which participants would be asked, “Ever taken prednisone or cortisone daily?” However, osteoporosis induced by glucocorticoids was also associated with the duration or average dose of glucocorticoids use [36, 37]. Moreover, glucocorticoids included many different species and kinds of glucocorticoids except for prednisone and cortisone. Therefore, although we were unable to demonstrate our hypothesis because related data were not obtainable in the NHANES database, the reason above might be a plausible explanation for the exclusion of the four variables. Second, the association between the predictors and the risk of osteoporosis might be confounded by other confounders like age, sex, or other variables not included in the present study (such as postmenopausal status). Third, recall bias might affect the results because all data of the four variables were collected based on the

questionnaire or interview. Finally, some other reasons such as the small sample size of the included participants might be the potential influencing factors that influenced the association between the excluded variables and the risk of osteoporosis. In summary, the above contents provided a reference and pointed some caveats to consider for future studies.

According to the coefficient in the first analysis, participants who had less than 12 times drinking during the past year had a higher risk of OP than those who had at least 12 times drinking, which appeared to contradict previous conclusions that alcohol consumption was positively associated with the risk of OP [32, 33]. We considered the observed association between alcohol consumption and the risk of osteoporosis might be confounded by other variables. Actually, according to the analysis of the age and sex ratio for different alcohol consumption statuses, the results showed that participants who consumed less time tended to be older and more often female (Supplementary Table S4), which was a plausible explanation for this phenomenon. In addition, although the initial model showed that individuals with MVPA had a lower risk of OP than those with no MVPA, older adults with low MVPA appeared to have a lower risk of osteoporosis than those with medium or high MVPA, according to the related coefficient in the final model (Supplementary Table S3), which might mislead and give the wrong impression to users (especially non-professionals) that less MVPA could be beneficial to prevent osteoporosis. On the one hand, the association might be confounded by the other variables. On the other hand, the relatively small sample size of subjects with low MVPA (non-OP: $n = 325$ [18.2%]; OP: $n = 16$ [10.6%]) might also be a potential reason for this phenomenon. Therefore, to overcome this drawback, we modified the grouping criterion by dividing all participants into three groups (no MVPA, low or medium MVPA, and high MVPA), and the coefficients of the final models suggested that the risk of osteoporosis decreased with the increases in MVPA, which was in line with findings from the previous studies [34, 35]. Moreover, these results also suggest that more large-sample studies are needed to improve the risk prediction model.

Compared with other risk prediction models for OP, the model in this study also had some advantages. Wang et al. developed a prediction model for OP based on a neural network [10], which showed an AUC of 0.901 in the training set and 0.762 in the test set. More predictors ($n = 18$) and the difference between the model in the training set and the test set might be the limitations of this study. Ho-Pham et al. also developed a prediction model for OP based on logistic regression [11]. Although the model showed an AUC of 0.825 for women and 0.858 for men, external validation was not performed, which might lead to an overestimated model performance. In addition, Koh et al.

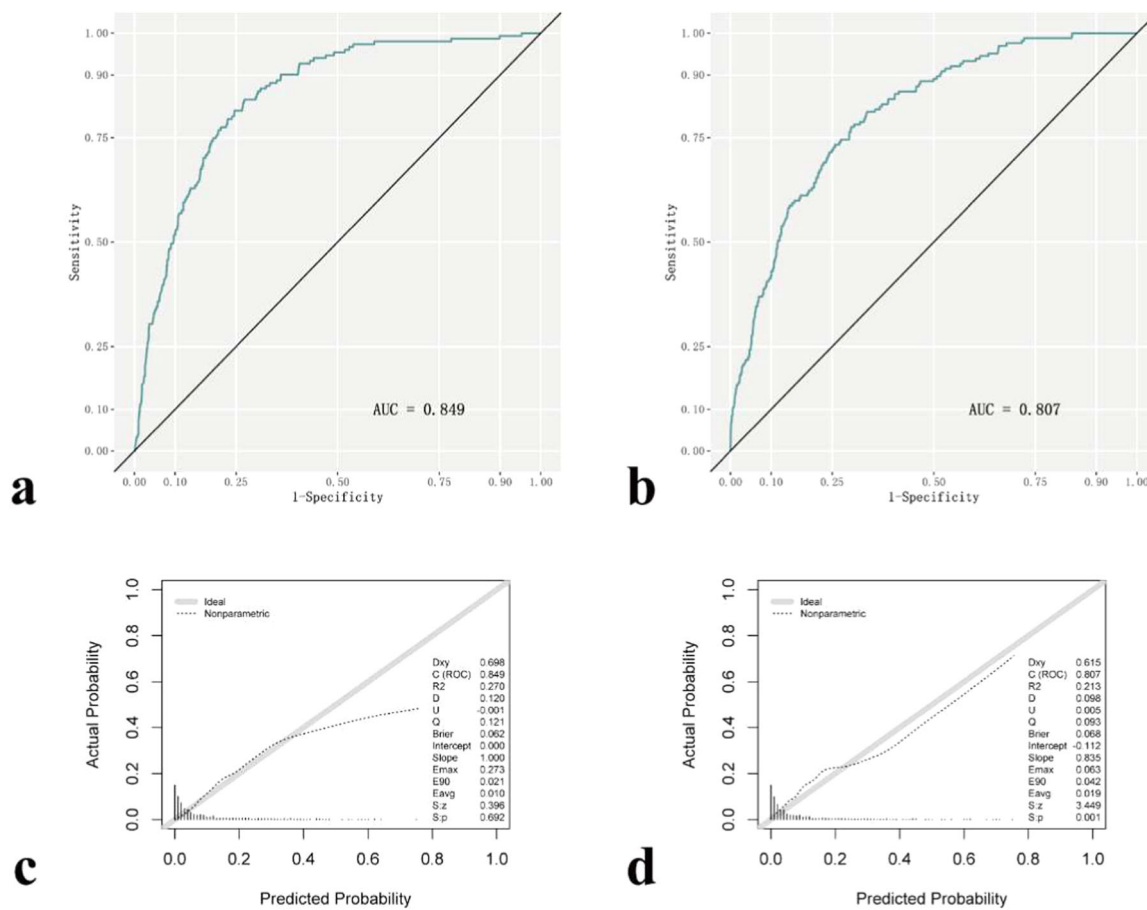


Fig. 2 The discrimination (AUC) and calibration (Brier score) of the screening model in the development cohort and validation cohort. **a, c** for development cohort; **b, d** for validation cohort. AUC Area under the receiver operator characteristics

Table 3 Model performance

	Apparent performance ^a	Test performance ^b	Average optimism ^c	Optimism corrected ^d	External validation performance
AUC	0.849 (0.820–0.878)	0.838 (0.837–0.838)	+0.017	0.833 (0.831–0.834)	0.807 (0.775–0.839)
Brier score	0.062 (0.054–0.070)	0.063 (0.063–0.063)	−0.002	0.064 (0.064–0.064)	0.068 (0.059–0.077)

^aRefers to performance estimated directly from dataset that was used to develop prediction model

^bDetermined by developing model in each bootstrap sample (500 samples with replacement), calculating performance (bootstrap performance), and applying bootstrap model in original sample

^cAverage difference between model performance in bootstrap data and test performance in original dataset

^dSubtracting average optimism from apparent performance

developed a simple tool (OSTA) to predict the risk of OP with 91% sensitivity and 45% specificity [12]. Ma et al. also developed a new predictive screening tool (BFH-OST) for screening OP with a sensitivity of 73.6% and specificity of 72.7% [13]. These two tools have advantages in use because only a small amount of simple information is required. However, these two tools were only applied for older women and performed relatively poorly compared to the model in this study (when using the cutoff of 0.072, the sensitivity and specificity were 0.848 and 0.732, respectively). In addition, the optimal cutoff value was determined

by the maximum Youden index (Youden index = sensitivity + specificity − 1) in the present study. It should be noted that the Youden index gives the same weight to sensitivity and specificity, but the sensitivity might be a more important indicator for the screening test because high sensitivity could reduce the rate of missed diagnosis. For example, if the researchers expect to reach a sensitivity of ≥ 90%, the threshold of 0.0492 might be an ideal cutoff value. However, there are cases where the specificity or both sensitivity and specificity played essential roles in diagnosis. Therefore, this study listed the local maxima of the ROC curve

Fig. 3 Screenshot of the model summary

Prediction model for osteoporosis

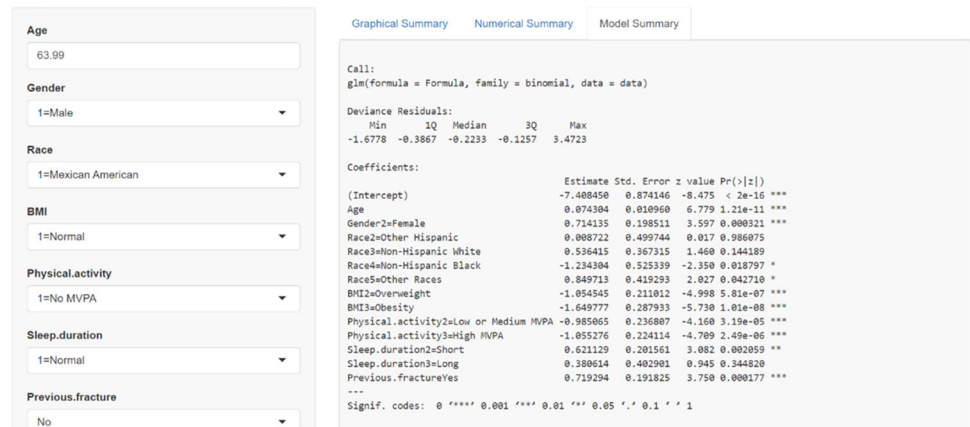
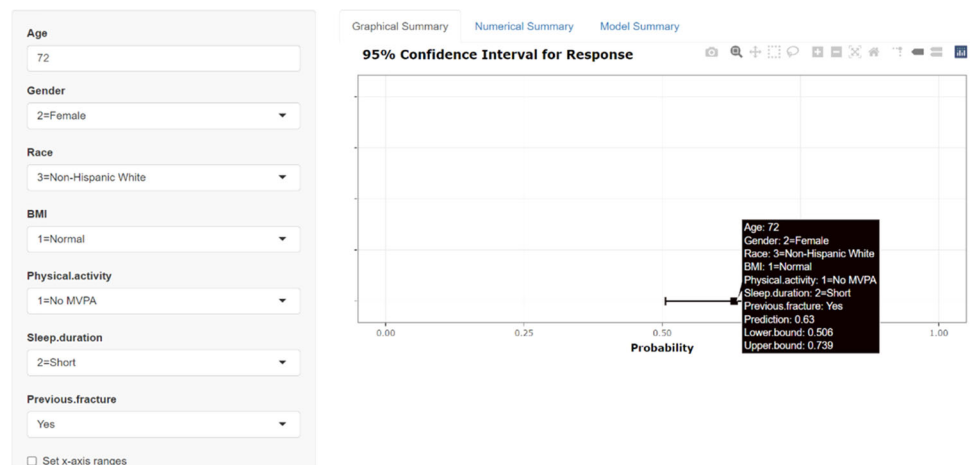


Fig. 4 Screenshot of an example for the results presentation



(Supplementary Table S5), including the specific values of threshold, sensitivity, specificity, positive predictive value, and negative predictive value and provided a reference for the tool use and future studies.

This risk prediction model has the potential to become a practical tool for the prevention of OP. First, based on the current status of the epidemiological data on OP, which showed a high incidence in older adults, it is necessary to develop a method for predicting the risk of OP. Second, all predictors in this model were based on the information obtained through questionnaires or interviews and do not require laboratory tests or physical examinations. This would be convenient to use, regardless of educational background. Third, the model performance showed excellent discrimination (C statistic above 0.800) and useful calibration (Brier scores below 0.070) in internal validation and external validation. Therefore, this prediction model also had relatively accurate predictive values. Finally, the gold standard for the diagnosis of OP is based on the BMD test, which is based on dual-energy X-ray absorptiometry and might not be optimal for OP screening. Therefore, an

interactive web application, which has advantages in comprehensibility and operationalization, appears to be a more practical method.

This prediction model also had some limitations as follows: (i) All data between the development and validation cohorts were based on individuals living in the United States. Therefore, more prospective, multicenter, and large-sample studies are needed to validate the performance of this prediction model. (ii) All predictors included in the final model were collected through interviews or questionnaires, which may have introduced recall bias and affected the accuracy of the model. (iii) Because the definition of OP was based on femur BMD information, this model may not be suitable for OP involving the spine or other sites. (iv) Although this model had relatively good performance for the diagnosis of OP, the number of predictors included in the final model was relatively large. Therefore, a prediction model for screening OP with high accuracy but a small number of predictors is worthy of further exploration in the future.

In conclusion, a novel risk prediction model for OP in older US adults, which has excellent discrimination and

good calibration, has been developed and externally validated. Considering its simplicity, generalizability, and accuracy, this risk prediction model has the potential to become a practical screening tool for OP.

Data availability

The datasets obtained and analysed during the current study are available on the NHANES database [<https://www.cdc.gov/nchs/nhanes/index.htm>].

Acknowledgements This study was supported by the National Natural Science Foundation of China (81874017, 81960403 and 82060405); Natural Science Foundation of Gansu Province of China (20JR5RA320); Cuiying Scientific and Technological Innovation Program of Lanzhou University Second Hospital (CY2017-ZD02); “Innovation Star” project for Excellent Graduate Students of the Education Department of Gansu Province (2021CXZX-143). At the same time, we would like to express our gratitude to Editage (<https://www.editage.cn/>) for the language editing services provided.

Author contributions Y.T. and Z.L. contributed equally to this work. Y.T.: conceptualization, methodology, investigation, resources, formal analysis, writing—original draft, writing—review & editing. Z.L.: conceptualization, methodology, investigation, resources, formal analysis, writing—review & editing. S.W.: methodology, investigation, writing—review & editing. Q.Y.: methodology, investigation, resources. Y.X.: writing—review & editing, funding acquisition. B.G.: conceptualization, methodology, writing—review & editing, funding acquisition.

Funding This study was supported by the National Natural Science Foundation of China (81874017, 81960403 and 82060405); Natural Science Foundation of Gansu Province of China (20JR5RA320); Cuiying Scientific and Technological Innovation Program of Lanzhou University Second Hospital (CY2017-ZD02); Innovation Star Project for Excellent Graduate Students of the Education Department of Gansu Province (2021CXZX-143).

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. All analyses were based on data of the National Health and Nutrition Examination Survey (NHANES). The study was approved by the ethics review board of the National Center for Health Statistics. The National Center for Health Statistics Ethics Review Board protocol numbers are Continuation of Protocol #2011-17 (NHANES 2013–2014 and 2017–2018), Protocol #2018-01 (NHANES 2017–2018), respectively. The detailed information located on the NHANES website.

Consent to participate Written informed consent was obtained from each participant before their inclusion on the NHANES database. Detailed information on the ethics application and written informed consent are provided on the NHANES website.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. J.E. Compston, M.R. McClung, W.D. Leslie, Osteoporosis. *Lancet* **393**, 364–376 (2019)
2. IOF. About Osteoporosis. <https://www.osteoporosis.foundation/patients/about-osteoporosis> (2021) Accessed July 1 2021
3. J.J. Chamberlain, A.S. Rhinehart, C.F. Shaefer Jr, A. Neuman, Diagnosis and management of diabetes: synopsis of the 2016 American diabetes association standards of medical care in diabetes. *Ann. Intern Med* **164**, 542–552 (2016)
4. P.K. Whelton, R.M. Carey, W.S. Aronow, D.E. Casey Jr, K.J. Collins, C. Dennison Himmelfarb, S.M. DePalma, S. Gidding, K. A. Jamerson, D.W. Jones et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: a report of the American College of Cardiology/American Heart Association Task Force on clinical practice guidelines. *Hypertension* **71**, e13–e115 (2018)
5. B.N.R. Jaja, G. Sapoznik, H.F. Lingsma, E. Macdonald, K.E. Thorpe, M. Mamdani, E.W. Steyerberg, A. Molyneux, A.L.O. Manoel, B. Schatlo et al. Development and validation of outcome prediction models for aneurysmal subarachnoid haemorrhage: the SAHIT multinational cohort study. *Bmj* **360**, j5745 (2018)
6. D. Vistisen, G.S. Andersen, C.S. Hansen, A. Hulman, J.E. Henriksen, H. Bech-Nielsen, M.E. Jørgensen, Prediction of first cardiovascular disease event in Type 1 diabetes mellitus: the steno Type 1 risk engine. *Circulation* **133**, 1058–1066 (2016)
7. J. Hippisley-Cox, C. Coupland, Predicting risk of osteoporotic fracture in men and women in England and Wales: prospective derivation and validation of QFractureScores. *Bmj* **339**, b4229 (2009)
8. J.A. Kanis, A. Oden, O. Johnell, H. Johansson, C. De Laet, J. Brown, P. Burckhardt, C. Cooper, C. Christiansen, S. Cummings 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* **18**, 1033–1046 (2007)
9. K.H. Rubin, S. Möller, T. Holmberg, M. Bliddal, J. Søndergaard, B. Abrahamsen, A. New, Fracture risk assessment tool (FREM) based on public health registries. *J. Bone Min. Res* **33**, 1967–1979 (2018)
10. Y. Wang, L. Wang, Y. Sun, M. Wu, Y. Ma, L. Yang, C. Meng, L. Zhong, M.A. Hossain, B. Peng, Prediction model for the risk of osteoporosis incorporating factors of disease history and living habits in physical examination of population in Chongqing, Southwest China: based on artificial neural network. *BMC Public Health* **21**, 991 (2021)
11. L.T. Ho-Pham, M.C. Doan, L.H. Van, T.V. Nguyen, Development of a model for identification of individuals with high risk of osteoporosis. *Arch. Osteoporos.* **15**, 111 (2020)
12. L.K. Koh, W.B. Sedrine, T.P. Torralba, A. Kung, S. Fujiwara, S. P. Chan, Q.R. Huang, R. Rajatanavin, K.S. Tsai, H.M. Park, J.Y. Reginster, A simple tool to identify asian women at increased risk of osteoporosis. *Osteoporos. Int* **12**, 699–705 (2001)
13. Z. Ma, Y. Yang, J. Lin, X. Zhang, Q. Meng, B. Wang, Q. Fei, BFH-OST, a new predictive screening tool for identifying osteoporosis in postmenopausal Han Chinese women. *Clin. Inter. Aging* **11**, 1051–1059 (2016)
14. S. Moradi, S. Shab-Bidar, S. Alizadeh, K. Djafarian, Association between sleep duration and osteoporosis risk in middle-aged and elderly women: A systematic review and meta-analysis of observational studies. *Metabolism* **69**, 199–206 (2017)

15. J. Xu, G. Lombardi, W. Jiao, G. Banfi, Effects of exercise on bone status in female subjects, from young girls to postmenopausal women: an overview of systematic reviews and meta-analyses. *Sports Med* **46**, 1165–1182 (2016)
16. R.S. Hardy, H. Zhou, M.J. Seibel, M.S. Cooper, Glucocorticoids and bone: consequences of endogenous and exogenous excess and replacement therapy. *Endocr. Rev.* **39**, 519–548 (2018)
17. CDC. National Health and Nutrition Survey. https://www.cdc.gov/nchs/nhanes/about_nhanes.htm (2021) Accessed July 1 2021
18. CDC. NCHS Research Ethics Review Board (ERB) Approval. <https://www.cdc.gov/nchs/nhanes/irba98.htm> (2021) Accessed July 1 2021
19. CDC. NHANES 2013–2014. <https://www.cdc.gov/nchs/nhanes/ContinuousNhanes/Default.aspx?BeginYear=2013> (2021) Accessed July 1 2021
20. CDC. NHANES 2017–2018. <https://www.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2017> (2021) Accessed July 1 2021
21. D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Bmj* **339**, b2535 (2009)
22. P. Peduzzi, J. Concato, E. Kemper, T.R. Holford, A.R. Feinstein, A simulation study of the number of events per variable in logistic regression analysis. *J. Clin. Epidemiol.* **49**, 1373–1379 (1996)
23. B.H. Huang, M.J. Duncan, P.A. Cistulli, N. Nassar, M. Hamer, E. Stamatakis, Sleep and physical activity in relation to all-cause, cardiovascular disease and cancer mortality risk. *Br J Sports Med.* (2021). Online ahead of print
24. WHO. Physical activity. <https://www.who.int/news-room/fact-sheets/detail/physical-activity> (2021) Accessed July 1 2021
25. CDC. Appendix I. Suggested MET Scores. https://www.cdc.gov/Nchs/Nhanes/2013-2014/PAQ_H.htm#Appendix_1. _Suggested_MET_Scores (2021) Accessed July 1 2021
26. M. Hirshkowitz, K. Whiton, S.M. Albert, C. Alessi, O. Bruni, L. DonCarlos, N. Hazen, J. Herman, E.S. Katz, L. Kheirandish-Gozal et al. National Sleep Foundation's sleep time duration recommendations: methodology and results summary. *Sleep. Health* **1**, 40–43 (2015)
27. CDC. Dual-Energy X-ray Absorptiometry (DEXA) Scan. https://www.cdc.gov/nchs/data/nhanes/2013-2014/manuals/2013_Body_Composition_DXA.pdf (2021) Accessed July 1 2021
28. A.C. Looker, E.S. Orwoll, C.C. Johnston Jr, R.L. Lindsay, H.W. Wahner, W.L. Dunn, M.S. Calvo, T.B. Harris, S.P. Heyse, Prevalence of low femoral bone density in older U.S. adults from NHANES III. *J. Bone Min. Res* **12**, 1761–1768 (1997)
29. R.R. Seethala, P.C. Hou, I.P. Aisiku, G. Frendl, P.K. Park, M.E. Mikkelsen, S.Y. Chang, O. Gajic, J. Sevransky, Early risk factors and the role of fluid administration in developing acute respiratory distress syndrome in septic patients. *Ann. Intensive Care* **7**, 11–11 (2017)
30. U.A. Tahir, G. Doros, J.S. Kim, L.H. Connors, D.C. Seldin, F. Sam, Predictors of mortality in light chain cardiac amyloidosis with heart failure. *Sci. Rep.* **9**, 8552–8552 (2019)
31. A. Chiu, M. Ayub, C. Dive, G. Brady, C.J. Miller, twodpccr: an R/Bioconductor package and Shiny app for Droplet Digital PCR analysis. *Bioinformatics* **33**, 2743–2745 (2017)
32. Z. Cheraghi, A. Doosti-Irani, A. Almasi-Hashiani, V. Baigi, N. Mansournia, M. Etminan, M.A. Mansournia, The effect of alcohol on osteoporosis: A systematic review and meta-analysis. *Drug Alcohol Depend.* **197**, 197–202 (2019)
33. K. Zhu, R.L. Prince, Lifestyle and osteoporosis. *Curr. Osteoporos. Rep.* **13**, 52–59 (2015)
34. K.T. Borer, Physical activity in the prevention and amelioration of osteoporosis in women: interaction of mechanical, hormonal and dietary factors. *Sports Med.* **35**, 779–830 (2005)
35. M.B. Pinheiro, J. Oliveira, A. Bauman, N. Fairhall, W. Kwok, C. Sherrington, Evidence on physical activity and osteoporosis prevention for people aged 65+ years: a systematic review to inform the WHO guidelines on physical activity and sedentary behaviour. *Int J. Behav. Nutr. Phys. Act.* **17**, 150 (2020)
36. P. Chotiyarnwong, E.V. McCloskey, Pathogenesis of glucocorticoid-induced osteoporosis and options for treatment. *Nat. Rev. Endocrinol.* **16**, 437–447 (2020)
37. L. Buckley, M.B. Humphrey, Glucocorticoid-induced osteoporosis. *N. Engl. J. Med* **379**, 2547–2556 (2018)