

Development of a Model for Predicting the 4-year Risk of Symptomatic Knee Osteoarthritis in China: A Longitudinal Cohort Study

Limin Wang

Peking University <https://orcid.org/0000-0001-7783-0748>

Han Lu

Peking University School of Nursing

Hongbo Chen

Peking University School of Public Health

Shida Jin

Peking University School of Nursing

Mengqi Wang

Peking University School and Hospital of Stomatology: Peking University School of Stomatology

Shaomei Shang (✉ shangshaomei@126.com)

Peking University, School of Nursing

Research article

Keywords: knee osteoarthritis, risk, prediction model

Posted Date: February 3rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-86446/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on February 26th, 2021. See the published version at <https://doi.org/10.1186/s13075-021-02447-5>.

Abstract

Objectives: We aimed to develop a model for predicting the 4-year risk of knee osteoarthritis (KOA) based on survey data obtained via a random, nationwide sample of Chinese individuals.

Methods: Data was analyzed from 8,193 middle-aged and older adults included in the China Health and Retirement Longitudinal Study (CHARLS). The incident of symptomatic KOA was defined as participants who were free of symptomatic KOA at baseline (CHARLS2011) and diagnosed with symptomatic KOA at the 4-year follow-up (CHARLS2015). The effects of potential predictors on the incident of KOA were estimated using logistic regression models and the final model was internally validated using the bootstrapping technique. Model performance was assessed based on discrimination—area under the receiver operating characteristic curve (AUC), and calibration.

Results: A total of 815 incidents of KOA were identified at the 4-year follow-up, resulting in a cumulative incidence of approximately 9.95%. The final multivariable model included age, sex, waist circumference, residential area, difficulty with activities of daily living (ADLs)/instrumental activities of daily living (IADLs), history of hip fracture, depressive symptoms, number of chronic comorbidities, self-rated health status, and level of moderate physical activity (MPA). The risk model showed good discrimination with AUC = 0.719 (95% confidence interval [CI] 0.700–0.737), optimism-corrected AUC = 0.712 after bootstrap validation. A satisfactory agreement was observed between the observed and predicted probability of incident symptomatic KOA. And a simple clinical score model was developed for quantifying the risk of KOA.

Conclusion: Our prediction model may aid the early identification of individuals at the greatest risk of developing KOA within 4 years.

Background

Knee osteoarthritis (KOA) is among the most common chronic diseases leading to disability worldwide, carrying a substantial and increasing health burden [1, 2]. The prevalence of symptomatic KOA and radiographic KOA in patients over 60 years of age ranges from 10.0–16.0% and 35.0–50.0% [3-7], respectively. Approximately 250 million people have KOA worldwide, with a two-fold increased prevalence in men and a three-fold increased prevalence in women in the United States over the past 20 years [5]; symptomatic KOA affects approximately 15.1 million individuals in the US population [8]. The estimated number of individuals over 60 years old suffering from symptomatic KOA reached 37.35 million in China [9]. The Years Lived with Disability (YLDs) caused by osteoarthritis ranked tenth in China in 2016 [10] and fourth in South Korea in 2015 [11]. Osteoarthritis had the fifth greatest relative increase in total YLDs from data of six Nordic countries from 1990 to 2015[12]. KOA was the first leading among osteoarthritis, accounting for 87% YLDs of osteoarthritis [13]. The increasing prevalence of KOA has increased the socioeconomic burden for affected individuals and healthcare systems [14].

To date, there are no effective therapeutic strategies for KOA. Prediction models for KOA aim to synthesize multiple factors to comprehensively predict the incident risk, and may allow for early detection and prevention [15]. The Nottingham KOA model was an early model for the prediction of 12-year KOA risk in middle-aged adults, including easily obtainable factors such as age, sex, family history, body mass index (BMI), occupational risk, and history of knee injury [16]; however, it was developed using data from only two communities in the UK, rather than a random sample of the general population, limiting its validity in other populations [17]. Several studies have developed prediction models based on genomic data [18-21] or radiographic/clinical biomarkers [22] such as hip α -angle and spinal bone mineral density. However, use of these models is limited due to their high cost or complexity [23].

Primary risk factors for incident KOA include advanced age, female gender, overweight/obesity, knee injury, and smoking [24-27]. Smoking decreases the risk of KOA, while the other factors increase the risk. Although physical activity [28, 29], occupational factors [24], ethnicity, and genetics [25] have also been associated with the incidence and/or progression of KOA, previous studies have reported inconsistent results due to methodological differences. Other potential risk factors for the development of KOA include metabolic syndrome [30-32], waist circumference [33], and depressive symptoms [24, 34], although findings regarding these factors remain controversial. Previous studies have reported a dual association between osteoarthritis and certain comorbidities (e.g., hypertension, ischemic heart disease, diabetes, etc.) [24, 25], suggesting that these comorbidities can influence the incidence and progression of KOA. Existing risk models of KOA have failed to include these potential risk factors, and there are currently no models for predicting KOA risk in the Chinese population. In this study, we aimed to develop a model for predicting the 4-year risk of KOA based on survey data obtained via a random, nationwide sample of Chinese individuals. This model would consider the potential risk factors.

Methods

Study design and data source

The present retrospective cohort study relied on 4-year data from the China Health and Retirement Longitudinal Study (CHARLS)—a nationwide study among Chinese adults aged 45 years or older for whom the detailed cohort profile has been published [35]. The national baseline survey for the study was conducted between June 2011 and March 2012 (CHARLS2011), and 17,708 respondents across 150 counties/districts and 450 villages/resident committees were recruited using a multistage sampling strategy. The respondents are followed up every 2 years via face-to-face computer-assisted personal interviews. Detailed information related to demographic background, socioeconomic status, biomedical findings, health status, and functioning was collected at baseline and at each follow-up using a structured questionnaire [35]. Blood samples were also obtained at each time point. The present study included participants recruited in CHARLS2011 and re-examined in CHARLS2015.

Participants

In this study, the unit of analysis was the person. Individuals who did not suffer from symptomatic KOA in CHARLS2011 and had complete diagnosis of symptomatic in CHARLS2015 were included. Participants who had no complete diagnosis of symptomatic KOA in CHARLS2011 or in CHARLS2015 were excluded. We also excluded those who had over 50% of predictive variables unavailable.

Outcomes

The primary outcome was the incident of symptomatic KOA during the 4-year follow-up period, and the subject was the unit of analysis. In accordance with the definition utilized in a previous study[36], symptomatic KOA was defined as both physician-diagnosed arthritis and the presence of concurrent pain in either knee joint. The incident of symptomatic KOA was defined as the participant being free of symptomatic KOA in CHARLS2011 and diagnosed with symptomatic KOA in CHARLS2015. The presence of pain in the knee joint was assessed based on responses to the following question: “Are you often troubled by pain in any part of your body?” If the participant answered in the affirmative, the following question was asked: “In what part of your body do you feel pain?”

Predictor variables

In CHARLS2011, data related to demographic background, socioeconomic status, biomedical findings, and levels of blood biomarkers was extracted. We included the following risk factors highlighted in previous studies and imputed missing values when necessary. The demographic variables included gender, age (year), BMI (categorized as underweight [$<18.5 \text{ kg/m}^2$], normal [$18.5\text{--}24.9 \text{ kg/m}^2$], overweight [$25.0\text{--}29.9 \text{ kg/m}^2$], obese [$\geq 30.0 \text{ kg/m}^2$]), waist circumference (cm) and residential area (urban vs. rural). Waist circumference was categorized into four groups: $<85/80 \text{ cm}$, $<90/85 \text{ cm}$, $<95/90 \text{ cm}$, and $\geq 95/90 \text{ cm}$ in men/women. The first group was referred to as the normal group and other three groups were referred to as central obesity based on the diagnosed criteria of central obesity recommended by the Department of Disease Control at the Ministry of Health [37]. The behavior variable included smoking status and engagement in vigorous/moderate/light physical activity. The physical activity score was calculated by multiplying the code for the duration by the code for frequency during one week [38]. According to physical activity score, the physical activity was divided into three levels (none: 0; low: 1–4; moderate-to-high ≥ 5). Health-related variables included history of hip fracture, number of other diagnosed comorbidities, metabolic syndrome (MS) in accordance with Chinese Diabetes Society (CDS) criteria [39], depressive symptoms based on Center for Epidemiologic Studies Depression Scale (CESD-10) score[40], self-rated health status and self-reported difficulties with activities of daily living (ADLs) [41] or instrumental activities of daily living (IADLs) [42]. The list of potential predictors is presented in Supplementary Box 1, along with detailed information related to how each predictor was assessed and the used tools.

Statistical analysis

Model structure

In CHARLS2011, physical activity measures were randomly available for 3,684 participants, while blood samples were available for 11,847 participants. Hence, physical activity scores of vigorous/moderate/light physical activity and MS were the main predictors with missing values. The percentage of missing values across the predictors varied between 0.04% and 57% in this study. We assumed data were missing at random and imputed 50 datasets based on the multiple imputation by chained equations (MICE) procedure [43]. The MICE technique improved the data accuracy, as any reasons for missing data could be explained by the observed variables included in the imputation model. We included all the predictor variables in the MICE process, along with the diagnosis of symptomatic KOA in CHARLS2011 and in CHARLS2015, as this information provides a stronger correlation structure among covariates used as predictors in the imputation model. Continuous variables (including systolic blood pressure, diastolic blood pressure, triacylglycerol, HDL cholesterol, and fasting blood glucose) were imputed using linear regression, and binary and multiple categorical variables (including duration and frequency of physical activity, history of hip fracture, smoking behavior, self-rated health status, CESD-10 items and ADL/IADLs items) were imputed using logit regression.

Descriptive statistics (means and standard deviations for continuous data, and counts and percentages for categorical data) were used to report key variables. Univariable and multivariable logistic regression analyses were used to establish a model for predicting the risk of KOA. All candidate variables were first evaluated via an unconditional univariable logistic regression analysis, and we then selected variables according to clinical value combined with statistical significance to conduct multivariable logistic regression analysis. In the multivariable logistic regression analysis, stepwise selection was combined with the Akaike information criterion (AIC) to determine the final model structure. The coefficients, odds ratios (ORs), and 95% CIs were estimated via 1,000-replication bootstrapping to obtain stable and unbiased parameters [44]. We combined the estimates using Rubin's rules [45].

Internal validation

The multivariable models were internally validated using a bootstrap procedure (sampling with replacement for 1,000 iterations) to assess bias-corrected estimates of predictive ability.

Model performance

We assessed the predictive performance of the final model using calibration and discrimination measures. Discrimination refers to the ability to distinguish patients experiencing an event from those not experiencing the event, and was quantified based on the area under the receiver operating characteristic curve (AUC) in this study. Calibration refers to how closely the predicted risk corresponds with the observed risk, and was assessed visually using calibration plots.

Clinical scoring tool

We developed a points-based risk-scoring tool based on the final model for easy clinical use—a widely utilized method of clinical scoring [23]. This clinical risk prediction tool can be used to identify individuals who are at high risk of developing KOA during the following 4 years. Continuous factors were categorized

based on the results of meta-analyses and clinical practice guidelines. Scores for categorical variables were determined by multiplying the b coefficients (log odds) in the multivariable logistic regression model by ten and rounding off decimal place. The total score was calculated by summing the scores of all variables. Sensitivity, specificity, and the AUC were calculated at different cut-off values, and the maximal Youden index was used to identify the optimal cut-off point [46]. The Youden index was calculated as follows: sensitivity + specificity – 1.

The present study was conducted in accordance with the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines for model development and reporting. All analyses were performed using STATA version 15.1 (STATA Corporation, College Station, TX) and R version 3.6.3 (R Foundation for Statistical Computing, Vienna, Austria). All statistical tests were two-sided and *P* values of <0.05 were considered statistically significant.

Ethics statement

Given that the present study is a secondary analysis of publicly available CHARLS data, the Medical Ethics Board Committee of Peking University granted the study an exemption from review.

Results

In CHARLS2011, physical activity measures were available for 3,684 participants, while blood samples were available for 11,847 participants. Complete KOA data were available in CHARLS2011 and CHARLS2015 for 9,204 of these participants. Seven participants were excluded because they declined to undergo body measurements assessments, rendering over 50% of participant's variables (including measurements of weight, height, waist circumference, assessments of depressive symptoms, physical activity, ADLs/IADLs or the blood biomarkers) inaccessible. Among them, one patient was diagnosed with KOA in 2011, and one developed KOA in 2015. Among the remaining 9,197 participants, an additional 1,004 were excluded because they were diagnosed with KOA at baseline (CHARLS2011). Thus, data from a total of 8,193 patients were included when developing the model. Among the 8,193 included participants, 815 developed symptomatic KOA in the following 4 years. The overall 4-year cumulative incidence of symptomatic KOA was 9.95%, with 7.62% and 13.77% in males and females respectively.

The mean age was 58.82 years (standard deviation (SD) ± 9.01 years), and 4,251 patients were female (51.89%). At baseline, 23.31% participants had difficulty with ADLs/IADLs, while 17.08% were diagnosed with metabolic syndrome, and 44.66% reported one or two chronic comorbidities. A history of hip fracture was reported by 252 (3.08%) participants. Other baseline characteristics are summarized in Table 1, along with the number of missing values for each variable.

Univariable and multivariable analysis

Table 2 shows the results of the univariable and multivariable analysis based on the imputed datasets. In the univariable analysis, age was identified as a risk factor for KOA, with the biggest difference occurring between the 60–64 and 65–69 age groups. Female sex, rural residence, history of hip fracture, ADL/IADL

difficulty, severe depressive symptoms, more chronic comorbidities, poor health status, and higher levels of MPA were significantly associated with an increased risk of developing KOA (all P values ≤ 0.01), while smoking was significantly associated with a decreased risk of developing KOA ($P \leq 0.01$). Although high BMI/waist circumference and metabolic syndrome were also positively associated with the incidence of KOA, these associations were not significant (all P values >0.05). Considering the important effects of metabolism and vigorous physical activity on the incident of KOA, we included metabolic syndrome and level of vigorous physical activity (VPA) in the multivariable logistic model, although the significance was not significant. As the clinical values of BMI and waist circumference are comparable, we selected waist circumference into the multivariable logistic regression given the relatively smaller P values.

The final prediction model included ten variables: age, sex, waist circumference, residential area, ADLs/IADLs difficulty, history of hip fracture, depressive symptoms, number of chronic comorbidities, health status, and level of MPA.

Model performance

The discrimination and calibration curves for the model are shown as Figures 1 (a) and (b), respectively. The final prediction model achieved acceptable discrimination, AUC = 0.719 (95% CI: 0.700–0.737), with optimism = 0.007 and bias-corrected AUC = 0.712 after bootstrap validation. The apparent observed line was quite close to the ideal line, while the bias-corrected line was slightly further from the ideal line than the observed line after the bootstrap procedure.

Figure legends

Figure 1. The discrimination and calibration curves of final risk model.

Figure 1 (a): ROC curve analysis for predicting symptomatic KOA when using age, sex, waist circumference, residential area, ADL/IADL difficulty, history of hip fracture, depressive symptoms, number of chronic comorbidities, health status, and level of MPA. The AUC 0.719 (95% CI 0.700-0.737), and optimism-corrected AUC was 0.712 after bootstrap validation.

Figure 1 (b): The calibration curve.

Area under the receiver characteristic curve, AUC.

Receiver operating characteristic curve, ROC

Clinical score model

We developed a simple clinical score model based on the ten variables included in the final multivariable model (Table 3). Total scores in this model range from 0 (lowest risk) to 51 (greatest risk). This clinical score model may aid in identifying patients at the greatest risk for developing KOA within the next 4 years. The AUC of the risk score model was 0.713 (95% CI 0.695–0.731), and the optimal cut-off, where patients with a score ≥ 20.5 were most likely to develop KOA in 4 years, was obtained from the maximal

Youden index. At the optimal cut-off, the sensitivity and specificity were 63.3% and 66.0%, respectively. Referring to the previous score model [22], the incident probability of KOA within 4 years was calculated by dividing the total risk score by 51 and multiplying by 100%.

Discussion

We developed and internally validated a model for predicting the 4-year risk of symptomatic KOA among the Chinese population, based on data from the CHARLS cohort. An easy-to-use clinical score model was developed to identify individuals' risk of developing KOA. The model included ten convenient and accessible variables, including age, sex, and waist circumference, which are most commonly included in previous KOA prediction models. Besides we also included the other controversial or new predictors of KOA, which were first time tested in risk model of KOA. To our knowledge, this is the first model for predicting KOA risk in the Chinese population, and our results suggest that this model can be used to aid in the prevention of KOA.

Older age was identified as a risk factor for KOA in our study; the most significant increase in risk was observed in the 60–69 years group. The cumulative incidence of symptomatic KOA gradually increased from 45 years of age, increasing rapidly after 55 years of age, peaking at approximately 65 years of age [47]. After 70 years of age, increases in the cumulative incidence of KOA were no longer significant [47]. Our findings, along with previous, highlight the need to prevent the incident of KOA in individuals between 45 and 70 years of age.

Obesity creates an abnormal loading environment for weight-bearing joints and may contribute to the pathogenesis of KOA [48]. Alternatively, the increased risk of KOA may be caused by the positive energy balance and metaflammation associated with obesity [49]. Although BMI has been illustrated as an important predictor of KOA [50], Wallace et al. (2019) [48] reported that increased abdomen size is associated with a greater risk of radiographic KOA than high BMI. Further studies are required to determine whether BMI, waist circumference, or metabolic syndrome comprehensively influences KOA risk due to mechaflammation and metaflammation. In this study, we analyzed the effects of BMI, waist circumference, and metabolic syndrome on KOA incident in the Chinese population. None of these three factors were a significant predictor of KOA incident; BMI had relatively low significance compared with waist circumference.

We analyzed the likelihood that the damage by BMI on joint tissues and pain symptoms would not reach a significant effect in the short term. Only the 12-year Nottingham KOA model investigated BMI related to symptomatic KOA [16], and the nine-year Rotterdam model [51] and four-year Chingford model [22] were predictive for radiographic KOA. BMI was also not included in final four-year Chingford model. Zheng & Chen [50] synthesized that BMI was a significant factor for incident KOA, but the diagnosis of KOA was radiographic KOA or severe KOA or replacement of KOA in 13 of 14 included studies. Among the eight studies with follow-up duration less than ten years, none focused on symptomatic KOA. Another possible reason was that abnormal waist circumference was much prevalent than abnormal BMI in the Chinese

population because body feature is prone to be small in the Asian race compared with the European or American population [52], thus waist circumference was much significant with incident KOA than BMI in this study. The results imply that the sensitivity of index of obesity might vary with race when evaluating the risk of KOA. Additional studies focusing on risk model of KOA are required to verify the significance of BMI with the incident of symptomatic KOA in Chinese population and other ethnic populations.

Another conventional risk factor for KOA was physical activity. In the present study, no significant association was observed between VPA/LPA and the incident of KOA; however, MPA positively predicted the incident of KOA. The reported associations between physical activity and the incident of KOA were inconsistent, resulting from a variation in assessment methods, activity categories or populations. Felson et al. [28] reported walking and other recreational activities did not increase the risk of OA in older adults. Results from the Chingford cohort demonstrated that physical activities related to work and sports increase the risk of osteophytes, while walking decreases the risk of osteophytes in middle-aged women [53], however, all effects were not statistically significant. Findings from the Framingham Heart Study [54] indicate that performing over 2 h per day would increase risk symptomatic KOA (OR: 5.3, 95% CI: 1.2–24) and the association was also significant for radiographic KOA (OR: 1.3 per hour, 95% CI: 1.1–1.6), while the effects of MPA and LPA were insignificant. Given the discrepancy between studies, additional studies should aim to verify the influence of different types of physical activity on the risk of KOA. Such studies should seek to determine the most appropriate type, duration, frequency, and intensity of physical activity for preventing KOA in different populations.

In our model, health-related variables are addressed and our findings provide evidence that these variables contribute essential values to the prediction of symptomatic KOA. Depressive symptoms, comorbidities, and history of hip fracture are psychologically and physiologically objective factors related to KOA. Self-rated health and difficulty with ADL/IADLs were mainly subjective, which was reflected in the patient's knowledge and ability to cope with disease.

Patients with KOA are prone to be comorbid with depression and other chronic comorbidities, and chronic diseases often exhibit interactions with comorbidities in complex ways [55]. Hence, previous studies have assumed that there may be a potential effect of depression and chronic comorbidities on incident of KOA. Seavey et al. [34] indicated that depressive symptoms represented a risk factor for arthritis incident (OR: 1.72, 95% CI: 1.27–2.35). Jinks et al. [56] also reported that depression was a significant predictor of knee pain (OR: 1.4, 95% CI: 1.1–1.8), where pain is the dominant physical symptom among patients with symptomatic KOA. Our study is the first model involving depression as predictor in a prediction model of symptomatic KOA. Patients with mild or moderate-to-severe depression were two or three times more likely to develop KOA than those without depression. Although a bidirectional causal association has rarely been illustrated either between arthritis and depression or between any other chronic disease and depression, targeted strategies for addressing depressive symptoms may therefore aid in reducing the incident of KOA. We also assessed relationships for 12 main types of comorbidities with incident of KOA. KOA and comorbidities may accelerate the progression of one another [24]. Results showed that patients with comorbidities had a significantly increased risk of developing KOA within four years. This addressed

the effect of comorbidities in developing KOA in the Chinese population; this might have some value for developing prediction model of KOA in other ethnic groups.

Related studies [57, 58] have demonstrated that rheumatoid arthritis increases the risk of hip fracture due to bone loss induced by chronic inflammation, use of glucocorticoids, and physical inactivity. However, rare studies indicated the association between hip fracture and KOA incident. Given that the knee and hip joints are the two most important weight-bearing joints, we sought to determine whether a history of hip fracture increases the risk of developing KOA. Our findings indicated that a history of hip fracture was associated with a 53% increase in the risk of KOA. Identifying the potential mechanisms underlying this association should be helpful for development of risk model in further studies.

Patient-reported outcome (PRO) has been emphasized in multiple studies because PROs, may capture important disease-related information prior to the onset of clinical signs or pathophysiological changes [59]. Silverwood et al. [24] noted that poor self-rated health status was a potential risk factor for KOA in an earlier study, although the association was insignificant. Our model showed that the likelihood of developing KOA increased as health status worsened, and impairments in ADLs/IADLs. Self-rated health status and assessments of difficulty with ADLs/IADLs could be significant predictors for incident of symptomatic KOA. Self-ratings of health status comprehensively reflect one's physical and psychological function, as well as one's knowledge and ability to cope with diseases and self-efficiency. Most of the existing potential risk factors were pooled from epidemiological analyses or clinicians' experience. Our findings highlight the need to consider the patient's perspective, as this may aid in furthering our understanding of KOA while reducing the incidence of the disease. Symptomatic KOA progressively decreases self-care ability, causing knee pain or stiffness. Our results implied that impairments in ADLs/IADLs prior to KOA onset may represent a predictive signal for KOA. Hence, preventive interventions may be useful in reducing the incident of KOA in those who have difficulty with ADL/IADL. Improving ADLs/IADLs might become a new interventional target to prevent KOA.

Preventing KOA or other chronic disease in rural area is the biggest challenge in China because of large population [36]. Our model included resident area as one predictor for KOA aiming to improve the prevention of KOA in rural population. Factors related to the high prevalence in rural areas may be multiple, including limited access to knowledge regarding the prevention of KOA and other chronic diseases, a lack of economic resources for timely treatment of chronic diseases, poor ability to manage one's health, and earlier impairments in physical function due to strenuous farm work. We hope that our results could promote policies and resources directed toward preventing KOA in Chinese rural areas in future.

We developed an easy-to-use clinical score model to identify risk of symptomatic KOA within 4 years and to identify individuals at high risk. This model involved ten commonly-available variables. This simple model involved ten commonly available variables, the assessment of variables, and the calculation of risk score, which are both easily understood and handled in practice. Clinicians or the patients themselves could use this tool to assess the risk of KOA in four-year term. While the score model showed a good

performance in assessing risk of KOA (AUC = 0.713; 95% CI 0.695–0.731), the ability to identify individuals at high risk was moderate using a -20.5 cut-off. Hence, this model should be improved and adjusted when applied in other populations. Adding other clinical biomarkers would provide further insight.

Limitations of study include, the incomplete data, though which was handled using the imputation method and a bootstrap strategy, may have biased our findings, especially the physical activity with a high percentage missing. Second, while new variables included in the study were significantly associated with the incident of KOA, further studies are required to elucidate the mechanisms underlying these associations. Lastly, our model was internally validated, therefore, external validation in other Chinese populations and different ethnic groups remains necessary.

Conclusion

In the present study, we developed the first model for predicting the 4-year risk of developing symptomatic KOA in China, using longitudinal cohort. Our simple score model may aid in the early identification of individuals at the greatest risk of developing KOA within 4 years in clinical practice or community setting. Such early identification may allow for improved patient education and modification of certain risk factors, which may in turn decrease rates of KOA incident.

Abbreviations

ADLs: Activities of daily living; AIC: Akaike information criterion; AUC: Area under the receiver operating characteristic curve; BMI: Body mass index; CESD: Center for Epidemiologic Studies Depression Scale; CHARLS: China Health and Retirement Longitudinal Study; CI: Confidence intervals; IADLs: Instrumental activities of daily living; KOA: Knee osteoarthritis; LPA: Light physical activity; MPA: Moderate physical activity; MS: Metabolic syndrome; OR: Odds ratio; PA: Physical activity; PRO: Patient-reported outcome; SD: Standard deviation; TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis; VPA: Vigorous physical activity.

Declarations

Ethics approval and consent to participate

The Medical Ethics Board Committee of Peking University granted the study an exemption from review.

Consent for publication

Not applicable.

Data availability statement

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (No. 81972158).

Author Contributions

All authors were involved in revising the article, and all authors approved the final version to be published. Study conception and design: LMW and SMS. Data analysis and interpretation: LMW, HL, HBC, SDJ, and MQW. Draft of the manuscript: LMW, HL, and SMS.

Acknowledgments

We thank the China Health and Retirement Longitudinal Study (CHARLS) team for providing nationally representative data.

References

1. Prieto-Alhambra D, Judge A, Javaid MK, Cooper C, Diez-Perez A, Arden NK. Incidence and risk factors for clinically diagnosed knee, hip and hand osteoarthritis: influences of age, gender and osteoarthritis affecting other joints. *Annals of the rheumatic diseases*. 2014;73(9):1659-64.
2. Hunter DJ, Schofield D, Callander E. The individual and socioeconomic impact of osteoarthritis. *Nature reviews Rheumatology*. 2014;10(7):437-41.
3. Cho HJ, Morey V, Kang JY, Kim KW, Kim TK. Prevalence and Risk Factors of Spine, Shoulder, Hand, Hip, and Knee Osteoarthritis in Community-dwelling Koreans Older Than Age 65 Years. *Clinical Orthopaedics and Related Research*. 2015;473(10):3307-14.
4. Lawrence RC, Felson DT, Helmick CG, Arnold LM, Choi H, Deyo RA, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. *ARTHRITIS AND RHEUMATISM*. 2008;58(1):26-35.
5. Nguyen U, Zhang YQ, Zhu YY, Niu JB, Zhang B, Felson DT. Increasing Prevalence of Knee Pain and Symptomatic Knee Osteoarthritis: Survey and Cohort Data. *ANNALS OF INTERNAL MEDICINE*. 2011;155(11):725-U135.
6. Postler A, Luque Ramos A, Goronzy J, Günther K-P, Lange T, Schmitt J, et al. Prevalence and treatment of hip and knee osteoarthritis in people aged 60 years or older in Germany: An analysis based on health insurance claims data. *Clinical Interventions in Aging*. 2018;13:2339-49.

7. Guillemin F, Rat AC, Mazieres B, Pouchot J, Fautrel B, Euller-Ziegler L, et al. Prevalence of symptomatic hip and knee osteoarthritis: a two-phase population-based survey. *Osteoarthritis Cartilage*. 2011;19(11):1314-22.
8. Deshpande BR, Katz JN, Solomon DH, Yelin EH, Hunter DJ, Messier SP, et al. The number of persons with symptomatic knee osteoarthritis in the United States: Impact of race/ethnicity, age, sex, and obesity. *Arthritis care & research*. 2016;68(12):1743-50.
9. WANG L, CHEN H, LU H, YUE W, SHANG S. Research progress on disease burden and disease risk models of knee osteoarthritis. *Chinese Nursing Research*. 2020;34(20):3642-6.
10. ZENG X, QI J, YIN P, WANG L, LIU Y, LIU J, et al. 1990 to 2016 Disease burden in China and administrative regions. *Chinese Circulation Journal*. 2018;33(12):1147-58.
11. Radnaabaatar M, Kim Y-E, Go D-S, Jung Y, Yoon S-J. Burden of Disease in Coastal Areas of South Korea: An Assessment Using Health Insurance Claim Data. *International journal of environmental research and public health*. 2019;16(17):3044.
12. Kiadaliri AA, Lohmander LS, Moradi-Lakeh M, Petersson IF, Englund M. High and rising burden of hip and knee osteoarthritis in the Nordic region, 1990–2015: Findings from the Global Burden of Disease Study 2015. *Acta Orthopaedica*. 2018;89(2):177-83.
13. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet (London, England)*. 2018;392(10159):1789-858.
14. Abbott JH, Usiskin IM, Wilson R, Hansen P, Losina E. The quality-of-life burden of knee osteoarthritis in New Zealand adults: A model-based evaluation. *PLoS One*. 2017;12(10):e0185676.
15. Lee Y-h, Bang H, Kim DJ. How to Establish Clinical Prediction Models. *Endocrinology and metabolism (Seoul)*. 2016;31(1):38-44.
16. Zhang W, McWilliams DF, Ingham SL, Doherty SA, Muthuri S, Muir KR, et al. Nottingham knee osteoarthritis risk prediction models. *Ann Rheum Dis*. 2011;70(9):1599-604.
17. Michl GL, Katz JN, Losina E. Risk and risk perception of knee osteoarthritis in the US: a population-based study. *Osteoarthritis Cartilage*. 2016;24(4):593-6.
18. Valdes AM, Doherty M, Spector TD. The additive effect of individual genes in predicting risk of knee osteoarthritis. *Annals of the Rheumatic Diseases*. 2008;67(1):124-7.
19. Takahashi H, Nakajima M, Ozaki K, Tanaka T, Kamatani N, Ikegawa S. Prediction model for knee osteoarthritis based on genetic and clinical information. *Arthritis Research and Therapy*. 2010;12(5):R187-R.
20. Blanco FJ, Möller I, Romera M, Rozadilla A, Sánchez-Lázaro JA, Rodríguez A, et al. Improved prediction of knee osteoarthritis progression by genetic polymorphisms: the Arthrotest Study. *Rheumatology (Oxford, England)*. 2015;54(7):1236-43.
21. Kerkhof HJM, Bierma-Zeinstra SMA, Arden NK, Metrustry S, Castano-Betancourt M, Hart DJ, et al. Prediction model for knee osteoarthritis incidence, including clinical, genetic and biochemical risk factors. *Annals of the Rheumatic Diseases*. 2014;73(12):2116-21.

22. Garriga C, Sanchez-Santos MT, Judge A, Hart D, Spector T, Cooper C, et al. Predicting incident radiographic knee osteoarthritis in middle-aged women within 4 years: the importance of knee-level prognostic factors. *Arthritis Care Res (Hoboken)*. 2019.
23. Jiang WH, Wang JY, Shen XF, Lu WL, Wang Y, Li W, et al. Establishment and Validation of a Risk Prediction Model for Early Diabetic Kidney Disease Based on a Systematic Review and Meta-Analysis of 20 Cohorts. *DIABETES CARE*. 2020;43(4):925-33.
24. Silverwood V, Blagojevic-Bucknall M, Jinks C, Jordan JL, Protheroe J, Jordan KP. Current evidence on risk factors for knee osteoarthritis in older adults: a systematic review and meta-analysis. *Osteoarthritis and Cartilage*. 2015;23(4):507-15.
25. Glyn-Jones S, Palmer AJR, Agricola R, Price AJ, Vincent TL, Weinans H, et al. Osteoarthritis. *The Lancet (British edition)*. 2015;386(9991):376-87.
26. Kong L, Wang L, Meng F, Cao J, Shen Y. Association between smoking and risk of knee osteoarthritis: a systematic review and meta-analysis. *Osteoarthritis and Cartilage*. 2017;25(6):809-16.
27. Hui M, Doherty M, Zhang W. Does smoking protect against osteoarthritis? Meta-analysis of observational studies. *Annals of the Rheumatic Diseases*. 2011;70(7):1231-7.
28. Felson DT, Niu J, Clancy M, Sack B, Aliabadi P, Zhang Y. Effect of recreational physical activities on the development of knee osteoarthritis in older adults of different weights: The Framingham Study. *Arthritis Care & Research*. 2007;57(1):6-12.
29. Regnaud J-P, Regnaud J-P, Lefevre-Colau M-M, Lefevre-Colau M-M, Trinquart L, Trinquart L, et al. High-intensity versus low-intensity physical activity or exercise in people with hip or knee osteoarthritis. *The Cochrane database of systematic reviews*. 2015;2015;10(10):CD010203.
30. Courties A, Sellam J, Berenbaum F. Metabolic syndrome-associated osteoarthritis. *Current opinion in rheumatology*. 2017;29(2):214-22.
31. Li S, Felson DT. What Is the Evidence to Support the Association Between Metabolic Syndrome and Osteoarthritis? A Systematic Review. *Arthritis care & research (2010)*. 2019;71(7):875-84.
32. Maddah S, Mahdizadeh J. Association of metabolic syndrome and its components with knee osteoarthritis. *Acta Medica Iranica*. 2015;53(12):743-8.
33. Gill SV, Hicks GE, Zhang Y, Niu J, Apovian CM, White DK. The association of waist circumference with walking difficulty among adults with or at risk of knee osteoarthritis: the Osteoarthritis Initiative. *Osteoarthritis and Cartilage*. 2017;25(1):60-6.
34. Seavey WG, Kurata JH, Cohen RD. Risk factors for incident self-reported arthritis in a 20 year followup of the Alameda County Study Cohort. *Journal of rheumatology*. 2003;30(10):2103-11.
35. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort Profile: The China Health and Retirement Longitudinal Study (CHARLS). 2014.
36. Tang X, Wang S, Zhan S, Niu J, Tao K, Zhang Y, et al. The Prevalence of Symptomatic Knee Osteoarthritis in China: Results From the China Health and Retirement Longitudinal Study. *Arthritis & rheumatology (Hoboken, NJ)*. 2016;68(3):648-53.

37. Chen C, Lu FC, Department of Disease Control Ministry of Health PRC. The guidelines for prevention and control of overweight and obesity in Chinese adults. *Biomed Environ Sci.* 2004;17:1-36.
38. MacInnis RJ, English DR, Hopper JL, Haydon AM, Gertig DM, Giles GG. Body Size and Composition and Colon Cancer Risk in Men. *Cancer Epidemiology Biomarkers & Prevention.* 2004;13(4):553-9.
39. Metabolic Syndrome Research Group of Chinese Medical Association. (2004). Suggestions on Metabolic Syndrome from Chinese Diabetes Society. *Chinese Journal of Diabetes Mellitus*, 12(03): p. 5-10.
40. Cheng S-T, Chan ACM. The Center for Epidemiologic Studies Depression Scale in older Chinese: thresholds for long and short forms. *International journal of geriatric psychiatry.* 2005;20(5):465-70.
41. Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. STUDIES OF ILLNESS IN THE AGED. THE INDEX OF ADL: A STANDARDIZED MEASURE OF BIOLOGICAL AND PSYCHOSOCIAL FUNCTION. *JAMA : the journal of the American Medical Association.* 1963;185:914-9.
42. Lawton MP, Lawton MP, Brody EM, Brody EM. Assessment of older people: Self-maintaining and instrumental activities of daily living. *Gerontologist.* 1969;9(3):179-86.
43. Schafer JL. Multiple imputation: a primer. *Statistical methods in medical research.* 2016;8(1):3-15.
44. Vergouwe Y, Royston P, Moons KGM, Altman DG. Development and validation of a prediction model with missing predictor data: a practical approach. *Journal of Clinical Epidemiology.* 2010;63(2):205-14.
45. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in medicine.* 2011;30(4):377-99.
46. Zhang L, Guo L, Wu H, Gong X, Lv J, Yang Y. Role of physical performance measures for identifying functional disability among Chinese older adults: Data from the China Health and Retirement Longitudinal Study. *PloS one.* 2019;14(4):e0215693.
47. Losina E, Weinstein AM, Reichmann WM, Burbine SA, Solomon DH, Daigle ME, et al. Lifetime Risk and Age at Diagnosis of Symptomatic Knee Osteoarthritis in the US. *Arthritis care & research (2010).* 2013;65(5):703-11.
48. Wallace IJ, Felson DT, Worthington S, Duryea J, Clancy M, Aliabadi P, et al. Knee osteoarthritis risk in non-industrial societies undergoing an energy balance transition: evidence from the indigenous Tarahumara of Mexico. *Annals of the Rheumatic Diseases.* 2019;78(12):1693-8.
49. Berenbaum F, Wallace IJ, Lieberman DE, Felson DT. Modern-day environmental factors in the pathogenesis of osteoarthritis. *Nature reviews Rheumatology.* 2018;14(11):674-81.
50. Zheng H, Chen C. Body mass index and risk of knee osteoarthritis: systematic review and meta-analysis of prospective studies. *BMJ Open.* 2015;5(12):e007568.
51. Kerkhof HJ, Bierma-Zeinstra SM, Arden NK, Metrustry S, Castano-Betancourt M, Hart DJ, et al. Prediction model for knee osteoarthritis incidence, including clinical, genetic and biochemical risk factors. *Ann Rheum Dis.* 2014;73(12):2116-21.

52. Association MSRGoCM. Suggestions on Metabolic Syndrome from Chinese Diabetes Society. Chinese Journal of Diabetes Mellitus. 2004(03):5-10.
53. Hart DJ, Doyle DV, Spector TD. Incidence and risk factors for radiographic knee osteoarthritis in middle-aged women: The Chingford Study. Arthritis & Rheumatism. 1999;42(1):17-24.
54. McAlindon TE, Wilson PWF, Aliabadi P, Weissman B, Felson DT. Level of physical activity and the risk of radiographic and symptomatic knee osteoarthritis in the elderly: the Framingham Study. The American journal of medicine. 1999;106(2):151-7.
55. Xu X, Mishra GD, Jones M. Evidence on multimorbidity from definition to intervention: An overview of systematic reviews. Ageing research reviews. 2017;37:53-68.
56. Jinks C, Jordan KP, Blagojevic M, Croft P. Predictors of onset and progression of knee pain in adults living in the community. A prospective study. Rheumatology (Oxford, England). 2007;47(3):368-74.
57. Arai K, Suzuki N, Murayama T, Kondo N, Otsuka H, Koizumi M, et al. Age at the time of hip fracture in patients with rheumatoid arthritis is 4 years greater than it was 10 years before, but is still younger than that of the general population. Modern rheumatology. 2019;30(1):64-9.
58. Yamamoto Y, Turkiewicz A, Wingstrand H, Englund M. Fragility fractures in patients with rheumatoid arthritis and osteoarthritis compared with the general population. Journal of Rheumatology. 2015;42(11):2055-8.
59. Nelson EC, Eftimovska E, Lind C, Hager A, Wasson JH, Lindblad S. Patient reported outcome measures in practice. BMJ : British Medical Journal. 2015;350(feb10 14):g7818-g.

Tables

Table 1 Baseline characteristics and outcomes of the study cohort summarized by their count and fraction (N (%)) for categorical or the mean and standard deviation for continuous variables, respectively

Characteristics	Number imputed points	of data (N=8193)	No symptomatic KOA(N=7378)	Symptomatic KOA(N=815)	P
mean (SD)	0	58.82(9.01)	58.73(9.06)	59.60(8.45)	<0.01
le	0	3,942 (48.11)	3,664(49.66)	278 (34.11)	<0.01
nale		4,251 (51.89)	3,714 (50.34)	537 (65.89)	
mean (SD)	0	23.46 (3.49)	23.44(3.47)	23.47(3.48)	0.805
2	0	85.31(9.58)	85.34(9.54)	85.12(9.56)	0.521
t ference, l (SD) cm	0				<0.01
lence Area	0				<0.01
ral, %		5,406(65.98)	4816(65.28)	590(72.39)	
ban, %		2,787(34.02)	2562(34.72)	225(27.61)	
ce Behavior	40				<0.01
smoking, %		4,945(60.36)	4374(59.28)	571(70.06)	
smoking, %		687(8.39)	624(8.46)	63(7.73)	
urrent smoking,		2,561(31.26)	2380(32.26)	181(22.21)	
IADL	16				<0.01
ulty					
, %		6,283(76.69)	5787(78.44)	496(60.86)	
, %		1,910(23.31)	1591(21.56)	319(39.14)	
	1312				0.708
, %		6,794(82.92)	6122(82.98)	672(82.45)	
, %		1,399(17.08)	1256(17.02)	143(17.55)	
racture	72				<0.01
, %		7,941(96.92)	7167(97.14)	774(94.97)	
, %		252(3.08)	211(2.86)	41(5.03)	
ession	1427				<0.01
, %		6,132(74.84)	5702(77.28)	430(52.76)	
d, %		1,896(23.14)	1561(21.16)	335(41.1)	
Moderate-to-		165(2.01)	115(1.56)	50(6.13)	
e, %					
orbidities	0				<0.01
ne, %		3,778(46.11)	3498(47.41)	280(34.36)	
2, %		3,659(44.66)	3253(44.09)	406(49.82)	
%		756(9.23)	627(8.5)	129(15.83)	

Characteristics	Number imputed points	of data (N=8193)	No symptomatic KOA(N=7378)	Symptomatic KOA(N=815)	<i>P</i>
Health Status	3				<0.01
Very good, %		289(3.53)	278(3.77)	8(0.98)	
Good, %		1,075(13.12)	1037(14.06)	38(4.66)	
Fair, %		2,752(33.59)	2538(34.4)	214(26.26)	
Poor, %		2,957(36.09)	2607(35.33)	350(42.94)	
Very poor, %		1,120(13.67)	915(12.4)	205(25.15)	
Physical Activity					
Metabolic score	4743				0.507
Low PA, %		4,942(60.32)	4462(60.48)	480(58.9)	
Low level, %		590(7.20)	524(7.1)	66(8.1)	
Middle-to-high %		2,661(32.48)	2392(32.42)	269(33.01)	
Metabolic score	4750				<0.01
Low PA, %		3,192(38.96)	2914(39.5)	278(34.11)	
Low level, %		891(10.88)	798(10.82)	93(11.41)	
Middle-to-high %		4,110(50.16)	3666(49.69)	444(54.48)	
Metabolic score	4751				0.842
Low PA, %		1,586(19.36)	1431(19.4)	155(19.02)	
Low level, %		1,389(16.95)	1245(16.87)	144(17.67)	
Middle-to-high %		5,218(63.69)	4702(63.73)	516(63.31)	

ADL = Activities of daily living; BMI = Body mass index; IADL = Instrumental activities of daily living; LPA = Light physical activity; MPA = Moderate physical activity; MS = Metabolic syndrome; PA = Physical activity; SD = Standard deviation; VPA = Vigorous physical activity.

Table 2 Results of Logistic Regression Models for Incident of KOA Generated from 8193 Participants in CHARLS 2011-2015

Predictors	Total N	KOA case n (%)	Univariable OR (95% CI)	Multivariable OR (95% CI)
Age (years)				
45-49	1,616	132(8.17)	Reference	Reference
50-54	1,241	112(9.02)	1.12(0.86-1.45)	1.06(0.81-1.39)
55-59	1,773	161(9.08)	1.12(0.88-1.43)	1.00(0.78-1.29)
60-64	1,477	183(12.39)	1.59(1.26-2.01) **	1.38(1.08-1.77) **
65-69	989	121(12.23)	1.57(1.21-2.03) **	1.31(1.00-1.72) **
≥70	1,097	106(9.66)	1.20(0.92-1.57)	0.95(0.71-1.26)
Sex				
Male	3,942	278(7.05)	Reference	Reference
Female	4,251	537(12.63)	1.91(1.64-2.22) **	1.51(1.22-1.87) ***
BMI level				
Normal weight	318	35(11.01)	Reference	—
Under weight	5,198	503(9.68)	1.15(0.80-1.66)	—
Overweight	2,201	225(10.22)	1.06(0.90-1.25)	—
Obese	476	52(10.92)	1.14(0.85-1.55)	—
Waist circumference, cm				
In male/female				
<85/80	3,313	307(9.27)	Reference	Reference
85~90/80~85	1,414	144(10.18)	1.11(0.90-1.37)	1.07(0.86-1.33)
90~95/85~90	1,348	141(10.46)	1.14(0.93-1.41)	1.06(0.85-1.33)
>95/90	2,118	223(10.53)	1.15(0.96-1.38)	1.06(0.86-1.32)
Residence Area				
Urban	2,787	225(8.07)	Reference	Reference
Rural	5,406	590(10.91)	1.39(1.19-1.64) **	1.24(1.05-1.47) *
Smoke Behavior				
No smoking	4,945	571(11.55)	Reference	Reference
Ex-smoking	687	63(9.17)	0.77(0.59-1.02)	0.98(0.71-1.35)
Current smoking	2,561	181(7.07)	0.58(0.49-0.69) **	0.85(0.68-1.08)
ADL/IADL Difficulty				
No	6,283	496(7.89)	Reference	Reference
Yes	1,910	319(16.7)	2.34(2.01-2.72) **	1.49(1.26-1.76) ***
MS				
No	6,794	672(9.89)	Reference	Reference
Yes	1,399	143(10.22)	1.13(0.68-1.89)	1.03(0.86-1.23)
Hip Fracture				
No	7,941	774(9.75)	Reference	Reference
Yes	252	41(16.27)	1.80(1.28-2.53) **	1.53(1.07-2.20) *
Depression				

No	6,132	430(7.01)	Reference	Reference
Mild	1,896	335(17.67)	2.85(2.44–3.32) **	2.04(1.73–2.41) ***
Moderate-to-severe	165	50(30.3)	5.77(4.08–8.15) **	3.34(2.32–4.82) ***
Comorbidities				
None	3,778	280(7.41)	Reference	Reference
1~2	3,659	406(11.1)	1.56(1.33–1.83) **	1.25(1.05–1.48) *
≥3	756	129(17.06)	2.57(2.05–3.22) **	1.54(1.19–1.98) **
Health Status				
Very good	286	8(2.8)	Reference	Reference
Good	1,075	38(3.53)	1.27(0.59–2.769)	1.13(0.52–2.46)
Fair	2,752	214(7.78)	2.93(1.43–6.00) **	2.21(1.07–4.54) *
Poor	2,957	350(11.84)	4.67(2.29–9.50) **	2.99(1.46–6.13) **
Very poor	1,120	205(18.3)	7.77(3.79–15.98) **	3.67(1.76–7.63) ***
Physical Activity Score				
VPA score				
No PA	4,942	480(9.71)	Reference	Reference
Low level	590	66(11.19)	1.17(0.89–1.54)	1.15(0.86–1.53)
Middle-to-high level	2,661	269(10.11)	1.05(0.89–1.22)	1.03(0.87–1.22)
MPA score				
No PA	3,192	278(8.71)	Reference	Reference
Low level	891	93(10.44)	1.22(0.95–1.56)	1.35(1.02–1.73) *
Middle-to-high level	4,110	444(10.8)	1.27(1.08–1.49) **	1.31 (1.10–1.56) **
LPA score				
No PA	1,586	155(9.77)	Reference	—
Low level	1,389	144(10.37)	1.07(0.84–1.36)	—
Middle-to-high	5,218	516(9.89)	1.01(0.84–1.22)	—

level,

*P≤0.05, **P≤0.01, ***P<0.001; ADL= Activities of daily living; BMI = Body mass index; 95% CI=95% Confidence interval; IADL= Instrumental activities of daily living; LPA = Light physical activity; MPA = Moderate physical activity; MS= Metabolic syndrome; OR= Odds ratio; PA = Physical activity; VPA = Vigorous physical activity.

Table 3 Risk score model of KOA incident prediction

Predictors	Risk score	Predictors	Risk score
Age (years)		Depression	
45-49	0	No	0
50-59	1	Mild	7
60-69	3	Moderate-to-severe	12
≥70	0	Comorbidities	
Sex		None	0
Male	0	1~2	2
Female	5	≥3	4
Waist circumference, cm		Health Status	
In male/female			
<85/80	0	Very good	0
≥85/80	1	Good	1
Residence Area		Fair	8
Urban	0	Poor	11
Rural	2	Very poor	13
ADLs/IADLs Difficulty		MPA score	
No	0	No PA	0
Yes	4	Low level	3
Hip Fracture		Middle-to-high level	3
No	0		
Yes	4		

Total score of 0 indicates lowest risk and score of 51 indicates greatest risk.

A cut-off of 20.5 was identified, which meant patients with a 20.5 or more total risk scores were at the greatest risk to develop symptomatic KOA in 4 years.

The incident probability of KOA within 4 years was calculated by the total risk score divided by 51 and multiplied by 100%.

MPA = Moderate physical activity; ADL= Activities of daily living; IADL= Instrumental activities of daily living.

Table 4
Risk score model of KOA incident prediction

Predictors	Risk score	Predictors	Risk score
Age (years)		Depression	
45–49	0	No	0
50–59	1	Mild	7
60–69	3	Moderate-to-severe	12
≥70	0	Comorbidities	
Sex		None	0
Male	0	1 ~ 2	2
Female	5	≥3	4
Waist circumference, cm, In male/female		Health Status	
< 85/80	0	Very good	0
≥85/80	1	Good	1
Residence Area		Fair	8
Urban	0	Poor	11
Rural	2	Very poor	13
ADL/IADL Difficulty		MPA score	
No	0	No PA	0
Yes	4	Low level	3
Hip Fracture		Middle-to-high level	3
No	0		
Yes	4		
Total score of 0 indicates lowest risk and score of 51 indicates greatest risk.			
A cutoff point of 20.5 was identified, which meant a patient with a 20.5 or more total risk scores had much risk to develop systematic KOA in 4 years, i.e., they were high risk individuals of KOA.			
The incident probability of KOA within 4 years was calculated by the total risk score divided by 51 and multiplied by 100%.			
MPA = Moderate physical activity; ADL = Activities of daily living; IADL = Instrumental activities of daily living.			

Figures

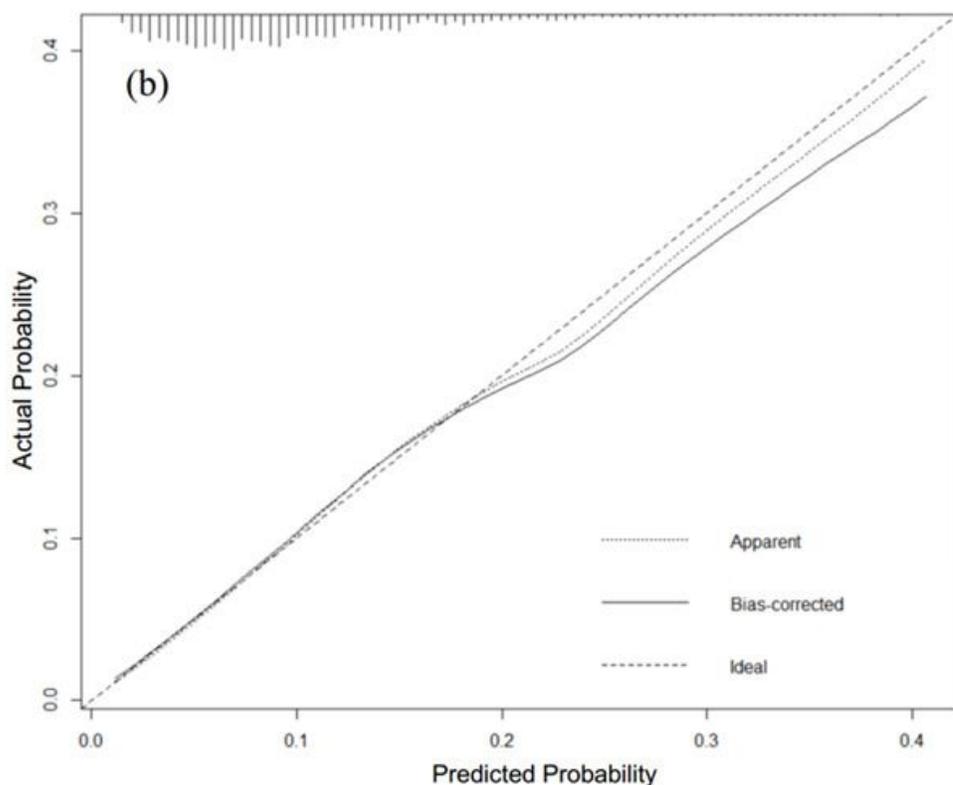
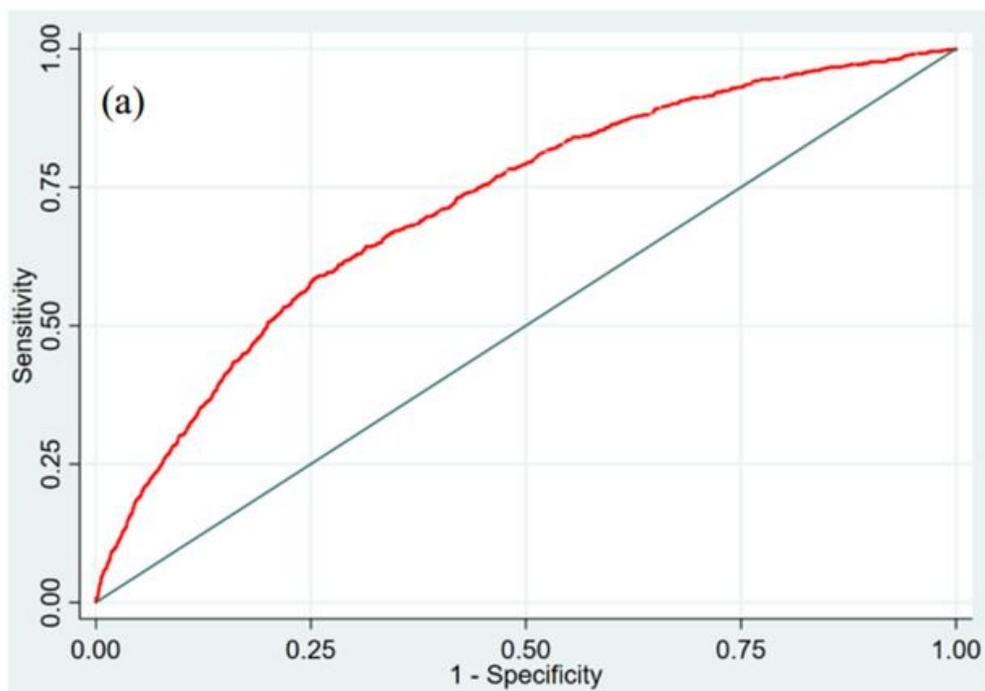


Figure 1

The discrimination and calibration curves of final risk model. (a): ROC curve analysis for predicting symptomatic KOA when using age, sex, waist circumference, residential area, ADL/IADL difficulty, history of hip fracture, depressive symptoms, number of chronic comorbidities, health status, and level of MPA.

The AUC 0.719 (95% CI 0.700-0.737), and optimism-corrected AUC was 0.712 after bootstrap validation.
(b): The calibration curve. Area under the receiver characteristic curve, AUC. Receiver operating characteristic curve, ROC

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AdditionalFile1.SBox1.Potentialpredictors.docx](#)
- [AdditionalFile2.SFigure1.ParticipantsFlowchart.docx](#)
- [AdditionalFile3.STable1.Baselinecharacteristic.docx](#)
- [AdditionalFile4.STable2.MultipleTestingResults.docx](#)