

Development of A Model for Predicting The 4-Year Risk of Systematic 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

Research article

Keywords: knee osteoarthritis, risk, prediction model

Posted Date: October 8th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-86446/v1>

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

Results: A total of 815 incident cases of KOA were identified at the four-year follow-up, resulting in a cumulative incidence of approximately 9.95%. The final multivariate 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 bias-corrected AUC for this model was 0.704. The calibration curve revealed satisfactory agreement between the observed and predicted incidence of systematic KOA. A simple clinical score model was developed for easily quantifying the risk of KOA based on these factors.

Conclusion: Our prediction model may aid in 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 systematic 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, and there are approximately 250 million people living with KOA worldwide. In the United States, the prevalence of KOA has increased twofold in men and threefold in women over the past 20 years [5], with systematic KOA affecting approximately 15.1 million individuals in the national population [8]. The number of patients with KOA in China is much higher due to its larger population and higher prevalence rates. The increasing prevalence of KOA has in turn increased the socioeconomic burden for affected individuals and healthcare systems.

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 for individuals and may allow for early detection and prevention [9]. The Nottingham KOA model was one 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 [10]. However, this model was

developed based on data from only two communities in North Nottinghamshire, rather than data from a random sample of the general population, limiting its validity in other populations [11]. Several studies have developed prediction models based on genomic data [12–15] or radiographic/clinical biomarkers [16] such as hip α -angle and spinal bone mineral density. However, these models are difficult to promote in clinical practice due to their high cost or complexity [17].

The primary risk factors for incident KOA include advanced age, female sex, overweight/obesity, knee injury, and smoking [18–21]. Among these, smoking decreases the risk of KOA, while the other factors increase the risk of KOA. Although physical activity [22, 23], occupational factors [18], ethnicity, and genetics [19] 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 [24–26], waist circumference [27], and depressive symptoms [18, 28], 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.) [18, 19], suggesting that these comorbidities can also influence the incidence and progression of KOA. However, 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. Therefore, in the present 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 and this model would consider these potential 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 [29]. 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 [29]. Blood samples were also obtained at each time point. The present study included participants recruited in CHARLS2011 and re-examined in CHARLS2015.

Outcomes

The primary outcome was the incident of systematic KOA during the 4-year follow-up period. In accordance with the definition utilized in a previous study, systematic KOA was defined as physician-diagnosed arthritis based on the presence of concurrent pain in the knee joint [30]. The incident of systematic KOA was defined as that the participants was free of systematic KOA in CHARLS2011 and

was diagnosed with systematic 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, we extracted data related to demographic background, socioeconomic status, biomedical findings, and levels of blood biomarkers. We included the following risk factors highlighted in previous studies and imputed missing values when necessary: sex (male vs. female), age (six groups: 45–49 years, 50–54 years, 55–59 years, 60–64 years, 65–69, ³70 years), smoking status (non-smoker, ex-smoker, current smoker), BMI (underweight (<18.5 kg/m²), normal (18.5–24.9 kg/m²), overweight (25.0–29.9 kg/m²), obese (\geq 30.0 kg/m²)), waist circumference (four categories in male/female participants: <85/80 cm, <90/85 cm, <95/90 cm, \geq 95/90 cm), residential area (urban vs. rural), hip fracture (yes vs. no), depressive symptoms based on Center for Epidemiologic Studies Depression Scale (CESD-10) score (none: \leq 11; mild: 12–23; moderate-to-severe: ³24), engagement in vigorous/moderate/light physical activity based on physical activity score (none: 0; low: 1–4; moderate-to-high \geq 5), number of diagnosed comorbidities (none, 1~ 2, ³3), and presence of metabolic syndrome (yes vs. no). We also included self-rated health status and self-reported difficulties with activities of daily living (ADLs) or instrumental activities of daily living (IADLs) as potential risk factors for KOA. The list of potential predictors is presented in Supplementary Box 1, along with detailed information related to how each predictor was assessed.

Statistical analysis

Model structure

We used multiple imputations based on five replications and a chained equation approach [31] to fill the missing values for systolic blood pressure, diastolic blood pressure, triacylglycerol, HDL cholesterol, fasting blood glucose, duration and frequency of physical activity, CESD-10 scores, and smoking status. We used descriptive statistics (means and standard deviations for continuous data and counts and percentages for categorical data) to report key variables.

We conducted univariable and multivariable logistic regression analysis to establish a model for predicting the risk of KOA. All candidate variables were first evaluated via an unconditional univariate logistic regression analysis, and we then selected variables according to the statistical significance and clinical value to conduct multivariate logistic regression analyses. In the multivariate 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% confidence intervals (CIs) were estimated via 1,000-replication bootstrapping to obtain stable and unbiased parameters [32].

Internal validation

The multivariate 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 [17]. This clinical risk prediction tool can be used to identify individuals who are high risk of developing KOA during the following 4 years. We categorized continuous factors 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 multivariate logistic regression model by 10 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 [33]. 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 <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 assessments of body measurements, making over 50% of variables 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 systematic KOA in the following 4 years. The prevalence of systematic KOA in CHARLS2011 and the 4-year cumulative incidence of systematic KOA in CHARLS2015 for the whole population and for different groups stratified by sex and age (≤60 years and >60 years) are shown in Table 1.

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 2.

Univariable and multivariable analysis

Table 3 shows the results of the univariate and multivariate analysis based on the imputed dataset, along with the number of missing values for each variable. 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/IDL difficulty, severe depressive symptoms, more chronic comorbidities, poor health status, and higher levels of MPA were associated with an increased risk of developing KOA, while smoking was associated with a decreased risk of developing KOA. Although high BMI/waist circumference and metabolic syndrome were also positively associated with the incidence of KOA, these associations were not significant. Considering the important effect of metabolism on the incident of KOA, we included waist circumference and metabolic syndrome in the multivariate model given their relatively smaller p-values.

The final prediction model included ten variables: 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.

Model performance

The discrimination and calibration curves for the model are shown as Figure 1a and Figure 1b, respectively. The final prediction model achieved acceptable discrimination (AUC = 0.718, optimism = 0.014, bias-corrected AUC = 0.704). The apparent observed line was quite close to the ideal line, while the bias-corrected line was slightly farther from the ideal line than the observed line after the bootstrap procedure.

Clinical score model

We developed a simple clinical score model based on the ten variables included in the final multivariate model (Table 4). 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, and the optimal cut-off where patients with a score greater than or equal to 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 [16], the incident probability of KOA within 4 years was calculated by dividing the total risk score by 51 and multiplying by 100%.

Discussion

In the present study, we developed and internally validated a model for predicting the 4-year risk of systematic KOA among the Chinese population, based on data from the CHARLS cohort. We also developed an easy-to-use clinical score model to identify individuals at the greatest risk of developing KOA. The final prediction model included ten convenient and accessible variables: age, sex, waist circumference, residential area, ADL/IADL difficulty, history of hip fracture, depressive symptoms, number of chronic comorbidities, self-reported health status, and level of MPA. Among these variables, age, sex, and waist circumference are the factors most commonly included in previous models for predicting KOA risk, while the others have been controversial or have not been considered in relation to KOA risk. To our knowledge, ours is the first model for predicting KOA risk in China, and our results suggest that this model can be used to aid in the prevention of KOA, thereby helping to reduce disease burden.

While 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. As estimated, the cumulative incidence of systematic KOA gradually increased beginning at 45 years of age, increasing rapidly after 55 years of age. The peak rate of increase was observed at approximately 65 years of age. However, after 70 years of age, increases in the cumulative incidence of KOA were no longer significant [34]. When taken with the results of previous studies, our findings highlight the need to prevent the incident of KOA in individuals from 45 to 70 years of age. Consistent with the results of previous studies [18, 19], female sex was also identified as a significant risk factor for the incident of KOA: after adjustment for other factors, the risk of developing KOA was 33% higher in women than in men.

Overweight/obesity increases the risk of KOA, in part because obesity creates an abnormal loading environment for weight-bearing joints. Obesity may also contribute to the pathogenesis of KOA by promoting chronic low-grade systemic inflammation or inducing mechanical damage to joint tissues, which in turn may lead to local inflammation. Together, these factors may weaken and degrade joint tissues [35]. Alternatively, the increased risk of KOA may be caused by the positive energy balance and metaflammation associated with obesity [36]. Metabolic syndrome has also been associated with the incident and progression of KOA [36]. Although BMI has been illustrated as an important predictor of KOA [37], Wallace et al. [35] 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 or waist circumference comprehensively influences KOA risk due to mechaflammation and metaflammation. In the present study, we separately analyzed the effect of BMI, waist circumference, and metabolic syndrome on KOA incident in the Chinese population using univariate logistic regression analyses. However, none of these three factors was identified as a significant predictor of KOA incident. While

similar insignificance also been reported for other ethnic groups [38], additional studies are required to verify this lack of significance in the Chinese population.

Our findings also indicated that depressive symptoms are a significant predictor of KOA risk. Patients with mild or moderate-to-severe depression were two or three times more likely to develop KOA than those without depression, respectively. Previous studies [39, 40] have revealed that KOA and other chronic diseases increase the odds of depression due to long-term treatment, chronic pain, or high treatment costs. Indeed, the pooled prevalence of depressive symptoms in patients with osteoarthritis has been reported as 19.9% [39]. Xu et al. demonstrated that depression, hypertension, diabetes, arthritis, asthma, and osteoarthritis are prone to be comorbid with other conditions [41]. In patients without any comorbidities at baseline, the risk of KOA remained higher in patients with depressive symptoms than in those without depressive symptoms. Adjusted ORs were 1.84 (95% CI: 1.39–2.43) and 3.29 (95% CI: 1.50–7.18) in patients with mild or moderate-to-severe depressive symptoms, respectively. Seavey et al. indicated that depressive symptoms represented a risk factor for arthritis incident based on data obtained from a 20-year longitudinal cohort study (OR: 1.72, 95% CI: 1.27–2.35) [28]. Jinks et al. also reported that depression was a significant predictor of knee pain (OR: 1.4, 95% CI: 1.1–1.8) while pain is the dominant physical symptom among patients with systematic KOA [42]. Although a bidirectional causal association has rarely been illustrated either between arthritis and depression or any other chronic disease and depression, depressive symptoms may be a significant risk factor for KOA. Targeted strategies for addressing depressive symptoms may therefore aid in reducing the incidence of KOA.

As patients with chronic diseases often exhibit comorbidities that interact with the disease in complex ways [41], we assessed relationships for 12 main types of comorbidities. The risk of developing KOA was 24% higher in patients with one or two comorbidities than in those without comorbidities. However, the risk of KOA increased by 54% in patients with three or more comorbidities. KOA and some comorbidities may accelerate the progression of one another [18], and comorbidities have been shown to predict worse pain and deteriorating physical function in patients with KOA [43]. The present study considered the effect of comorbidities when developing a prediction model for KOA in the Chinese population. Further studies including patients of different ethnic groups are required to determine whether this model can be applied more broadly.

Although physical activity is a modifiable risk factor for KOA, the reported associations between physical activity and outcomes have been inconsistent. In the present study, we observed no significant association between VPA/LPA and the incident of KOA; however, MPA positively predicted the incident of KOA. After adjustment for other factors, the risk of developing KOA was 30% greater in individuals engaging in all levels of MPA (low, middle, high) than in those engaging in no MPA. Notably, a previous study also reported that walking and other recreational activities did not increase the risk of OA in older adults [22]. Results from the Chingford cohort demonstrated that physical activity related to work (OR: 1.48, 95% CI: 0.34–5.64) and sports (OR: 1.23, 95% CI: 0.54–2.81) increases the risk of osteophytes, while physical activities such as walking decrease the risk of osteophytes (OR: 0.60, 95% CI: 0.22–1.71) in middle-aged women [44]. However, additional studies are required to verify these effects given that the

differences were not statistically significant. Findings from the Framingham Heart Study [45] indicate that long hours of heavy habitual physical activity per day are associated with an increased risk of KOA (OR: 1.3 per hour), with stronger effects occurring when strenuous physical activity is performed for more than 3 h per day (OR: 2.9, 95% CI: 1.2–6.9 for radiographic KOA and OR: 5.3, 95% CI: 1.2–24 for systematic KOA). In their study, the effects of MPA and LPA were insignificant. This discrepancy between our findings and previous results may be due in part to differences in the method of assessing physical activity levels or differences in the characteristics of the population. Given these differences, 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.

Assessments of difficulty in performing ADLs or IADLs may aid in the prediction of KOA incident. In the present study, the risk of developing KOA was 49.0% greater in individuals with ADL/IADL difficulty than in those without difficulty. Impairments in ADLs 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.

Our prediction model for KOA included one patient-reported outcome (PRO): self-rated health status. Numerous previous studies have focused on PROs, which may capture important disease-related information prior to the onset of clinical signs or pathophysiological changes [46]. Although Silverwood et al. noted that poor self-rated health status was a potential risk factor for KOA, no significant associations were observed [18]. In the present study, the odds of developing KOA increased as health status worsened. Compared to participants who rated their health status as “very good”, those who rated their health status as “fair” exhibited a 21% increase in the risk of KOA, while those who rated their health status as “poor” or “very poor” exhibited a 199% and 267% increase in the risk of KOA, respectively. 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. Although previous researchers have identified several potential risk factors, most of these factors were pooled from epidemiological analyses or clinical 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.

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. Related studies [47, 48] 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, studies reporting an association between hip fracture and KOA incidence are rare. Further studies are therefore required to determine whether hip fracture can be used to predict KOA risk, and to identify the potential mechanisms underlying this association.

In our study, the 4-year cumulative incidence of KOA in rural and urban areas was 10.9% and 8.1%, respectively. Residents of rural areas exhibited a 24% increase in the risk of developing KOA when compared with residents of urban areas, after controlling for other factors. Several previous studies have noted that symptomatic KOA is more common in rural residents of China than in urban residents [30, 49]. 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 may explain the increased risk of KOA in rural residents. These results suggest that policies and resources should be directed toward preventing KOA among residents of rural areas in China.

The present study possesses several limitations of note. First, although we used the imputation method to handle missing data and a bootstrap strategy to limit the influence of bias, incomplete data may still have biased our findings. Second, to efficiently predict the risk of KOA, we considered some new variables as predictors in our model. Though they were significantly associated with the incident of KOA, further studies are required to elucidate the mechanisms underlying these associations. Lastly, although we used a nationally representative cohort to develop the prediction model, the 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 systematic KOA in China, using CHARLS data. 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 incidence and progression.

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., et al., *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): p. 1659-1664.
2. Hunter, D.J., D. Schofield, and E. Callander, *The individual and socioeconomic impact of osteoarthritis*. *Nature reviews. Rheumatology*, 2014. **10**(7): p. 437-441.
3. Cho, H.J., et al., *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): p. 3307-3314.
4. Lawrence, R.C., et al., *Estimates of the prevalence of arthritis and other rheumatic conditions in the United States*. *ARTHRITIS AND RHEUMATISM*, 2008. **58**(1): p. 26-35.

5. Nguyen, U., et al., *Increasing Prevalence of Knee Pain and Symptomatic Knee Osteoarthritis: Survey and Cohort Data*. ANNALS OF INTERNAL MEDICINE, 2011. **155**(11): p. 725-U135.
6. Postler, A., 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**: p. 2339-2349.
7. Guillemin, F., et al., *Prevalence of symptomatic hip and knee osteoarthritis: a two-phase population-based survey*. Osteoarthritis Cartilage, 2011. **19**(11): p. 1314-22.
8. Deshpande, B.R., 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): p. 1743-1750.
9. Lee, Y.-h., H. Bang, and D.J. Kim, *How to Establish Clinical Prediction Models*. Endocrinology and metabolism (Seoul), 2016. **31**(1): p. 38-44.
10. Zhang, W., et al., *Nottingham knee osteoarthritis risk prediction models*. Ann Rheum Dis, 2011. **70**(9): p. 1599-604.
11. Michl, G.L., J.N. Katz, and E. Losina, *Risk and risk perception of knee osteoarthritis in the US: a population-based study*. Osteoarthritis Cartilage, 2016. **24**(4): p. 593-6.
12. Valdes, A.M., M. Doherty, and T.D. Spector, *The additive effect of individual genes in predicting risk of knee osteoarthritis*. Annals of the Rheumatic Diseases, 2008. **67**(1): p. 124-127.
13. Takahashi, H., et al., *Prediction model for knee osteoarthritis based on genetic and clinical information*. Arthritis Research and Therapy, 2010. **12**(5): p. R187-R187.
14. Blanco, F.J., et al., *Improved prediction of knee osteoarthritis progression by genetic polymorphisms: the Arthrotest Study*. Rheumatology (Oxford, England), 2015. **54**(7): p. 1236-1243.
15. Kerkhof, H.J.M., et al., *Prediction model for knee osteoarthritis incidence, including clinical, genetic and biochemical risk factors*. Annals of the Rheumatic Diseases, 2014. **73**(12): p. 2116-2121.
16. Garriga, 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.
17. Jiang, W.H., 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): p. 925-933.
18. Silverwood, V., et al., *Current evidence on risk factors for knee osteoarthritis in older adults: a systematic review and meta-analysis*. Osteoarthritis and Cartilage, 2015. **23**(4): p. 507-515.
19. Glyn-Jones, S., et al., *Osteoarthritis*. The Lancet (British edition), 2015. **386**(9991): p. 376-387.
20. Kong, L., et al., *Association between smoking and risk of knee osteoarthritis: a systematic review and meta-analysis*. Osteoarthritis and Cartilage, 2017. **25**(6): p. 809-816.
21. Hui, M., M. Doherty, and W. Zhang, *Does smoking protect against osteoarthritis? Meta-analysis of observational studies*. Annals of the Rheumatic Diseases, 2011. **70**(7): p. 1231-1237.

22. Felson, D.T., et al., *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): p. 6-12.
23. Regnaud, J.-P., 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): p. CD010203.
24. Courties, A., J. Sellam, and F. Berenbaum, *Metabolic syndrome-associated osteoarthritis*. Current opinion in rheumatology, 2017. **29**(2): p. 214-222.
25. Li, S. and D.T. Felson, *What Is the Evidence to Support the Association Between Metabolic Syndrome and Osteoarthritis? A Systematic Review*. Arthritis care & research (2010), 2019. **71**(7): p. 875-884.
26. Maddah, S. and J. Mahdizadeh, *Association of metabolic syndrome and its components with knee osteoarthritis*. Acta Medica Iranica, 2015. **53**(12): p. 743-748.
27. Gill, S.V., et al., *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): p. 60-66.
28. Seavey, W.G., J.H. Kurata, and R.D. Cohen, *Risk factors for incident self-reported arthritis in a 20 year followup of the Alameda County Study Cohort*. Journal of Rheumatology, 2003. **30**(10): p. 2103-2111.
29. Zhao, Y., et al., *Cohort Profile: The China Health and Retirement Longitudinal Study (CHARLS)*. 2014.
30. Tang, X., et al., *The Prevalence of Symptomatic Knee Osteoarthritis in China: Results From the China Health and Retirement Longitudinal Study*. Arthritis & rheumatology (Hoboken, N.J.), 2016. **68**(3): p. 648-653.
31. Schafer, J.L., *Multiple imputation: a primer*. Statistical methods in medical research, 2016. **8**(1): p. 3-15.
32. Vergouwe, Y., et al., *Development and validation of a prediction model with missing predictor data: a practical approach*. Journal of Clinical Epidemiology, 2010. **63**(2): p. 205-214.
33. Zhang, L., et al., *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): p. e0215693.
34. Losina, E., et al., *Lifetime Risk and Age at Diagnosis of Symptomatic Knee Osteoarthritis in the US*. Arthritis care & research (2010), 2013. **65**(5): p. 703-711.
35. Wallace, I.J., 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): p. 1693-1698.
36. Berenbaum, F., et al., *Modern-day environmental factors in the pathogenesis of osteoarthritis*. Nature reviews. Rheumatology, 2018. **14**(11): p. 674-681.

37. Zheng, H. and C. Chen, *Body mass index and risk of knee osteoarthritis: systematic review and meta-analysis of prospective studies*. *BMJ Open*, 2015. **5**(12): p. e007568.
38. Han, C.D., et al., *Correlation between metabolic syndrome and knee osteoarthritis: Data from the Korean National Health and Nutrition Examination Survey (KNHANES)*. *BMC Public Health*, 2013. **13**(1): p. 603-603.
39. Stubbs, B., et al., *Prevalence of depressive symptoms and anxiety in osteoarthritis: a systematic review and meta-analysis*. *Age and ageing*, 2016. **45**(2): p. 228-235.
40. Veronese, N., et al., *Association between lower limb osteoarthritis and incidence of depressive symptoms: data from the osteoarthritis initiative*. *Age and ageing*, 2017. **46**(3): p. 470-476.
41. Xu, X., G.D. Mishra, and M. Jones, *Evidence on multimorbidity from definition to intervention: An overview of systematic reviews*. *Ageing research reviews*, 2017. **37**: p. 53-68.
42. Jinks, C., et al., *Predictors of onset and progression of knee pain in adults living in the community. A prospective study*. *Rheumatology (Oxford, England)*, 2007. **47**(3): p. 368-374.
43. Calders, P. and A. Van Ginckel, *Presence of comorbidities and prognosis of clinical symptoms in knee and/or hip osteoarthritis: A systematic review and meta-analysis*. *Seminars in Arthritis and Rheumatism*, 2018. **47**(6): p. 805-813.
44. Hart, D.J., D.V. Doyle, and T.D. Spector, *Incidence and risk factors for radiographic knee osteoarthritis in middle-aged women: The Chingford Study*. *Arthritis & Rheumatism*, 1999. **42**(1): p. 17-24.
45. McAlindon, T.E., et al., *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): p. 151-157.
46. Nelson, E.C., et al., *Patient reported outcome measures in practice*. *BMJ : British Medical Journal*, 2015. **350**(feb10 14): p. g7818-g7818.
47. Arai, K., 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): p. 64-69.
48. Yamamoto, Y., et al., *Fragility fractures in patients with rheumatoid arthritis and osteoarthritis compared with the general population*. *Journal of Rheumatology*, 2015. **42**(11): p. 2055-2058.
49. Kang, X., et al., *The high prevalence of knee osteoarthritis in a rural Chinese population: The Wuchuan osteoarthritis study*. *Arthritis and rheumatism*, 2009. **61**(5): p. 641-647.

Tables

Table 1
The prevalence and four-year accumulative incidence of systematic KOA

Sex	Age (years)	Prevalence (%) § (n = 9197)	Four-year Accumulative Incidence (%)¶ (n = 8193)
Male	≤60	6.51	5.99
	>60	9.05	8.47
	Total	7.62	7.05
Female	≤60	12.12	11.48
	>60	16.58	14.69
	Total	13.77	12.63
§based on CHARLS2011.			
¶ based on CHARLS2015 and participants who were diagnosed systematic KOA in 2011 were excluded.			

Table 2

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 of imputed data points	Total (N = 8193)	No systematic KOA(N = 7378)	Systematic KOA(N = 815)	P
Age, mean (SD) years	0	58.82(9.01)	58.73(9.06)	59.60(8.45)	< 0.01
Sex	0				< 0.01
Male		3,942 (48.11)	3,664(49.66)	278 (34.11)	
Female		4,251 (51.89)	3,714 (50.34)	537 (65.89)	
BMI, mean (SD) kg/m ²	0	23.46 (3.49)	23.44(3.47)	23.47(3.48)	0.805
Waist circumference, mean (SD) cm	0	85.31(9.58)	85.34(9.54)	85.12(9.56)	0.521
Residence Area	0				< 0.01
Rural, %		5,406(65.98)	4816(65.28)	590(72.39)	
Urban, %		2,787(34.02)	2562(34.72)	225(27.61)	
Smoke Behavior	40			571(70.06)	< 0.01
No smoking, %		4,945(60.36)	4374(59.28)		
Ex-smoking, %		687(8.39)	624(8.46)	63(7.73)	
Current smoking, %		2,561(31.26)	2380(32.26)	181(22.21)	
ADL/IADL Difficulty	0				< 0.01
No, %		6,283(76.69)	5787(78.44)	496(60.86)	
Yes, %		1,910(23.31)	1591(21.56)	319(39.14)	
MS	1312				0.708
No, %		6,794(82.92)	6122(82.98)	672(82.45)	

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.

Characteristics	Number of imputed data points	Total (N = 8193)	No systematic KOA(N = 7378)	Systematic KOA(N = 815)	P
Yes, %		1,399(17.08)	1256(17.02)	143(17.55)	
Hip Fracture	72				< 0.01
No, %		7,941(96.92)	7167(97.14)	774(94.97)	
Yes, %		252(3.08)	211(2.86)	41(5.03)	
Depression	1427				< 0.01
No, %		6,132(74.84)	5702(77.28)	430(52.76)	
Mild, %		1,896(23.14)	1561(21.16)	335(41.1)	
Moderate-to-severe, %		165(2.01)	115(1.56)	50(6.13)	
Comorbidities	0				< 0.01
None, %		3,778(46.11)	3498(47.41)	280(34.36)	
1 ~ 2, %		3,659(44.66)	3253(44.09)	406(49.82)	
≥3, %		756(9.23)	627(8.5)	129(15.83)	
Health Status	3				< 0.01
Very good, %		286(3.49)	278(3.77)	8(0.98)	
Good, %		1,075(13.13)	1037(14.06)	38(4.66)	
Fair, %		2,752(33.60)	2538(34.4)	214(26.26)	
Poor, %		2,957(36.11)	2607(35.33)	350(42.94)	
Very poor, %		1,120(12.68)	915(12.4)	205(25.15)	
Physical Activity Score					0.507
VPA score	4743			480(58.9)	
No PA, %		4,942(60.32)	4462(60.48)	66(8.1)	
Low level, %		590(7.20)	524(7.1)	269(33.01)	

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.

Characteristics	Number of imputed data points	Total (N = 8193)	No systematic KOA(N = 7378)	Systematic KOA(N = 815)	<i>P</i>
Middle-to-high level, %		2,661(32.48)	2392(32.42)		
MPA score	4750				
No PA, %		3,192(38.96)	2914(39.5)	278(34.11)	< 0.05
Low level, % I		891(10.88)	798(10.82)	93(11.41)	
Middle-to-high level, %		4,110(50.16)	3666(49.69)	444(54.48)	
LPA score	4751				0.842
No 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 level, %		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 3
Results of Logistic Regression Models for Incident of KOA Generated from 8193 Participants in CHARLS
2011–2015

Predictors	Total N	KOA case n (%)	Univariate OR (95% CI)	Multivariate 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) **	1.33(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

*P≤0.05, **P < 0.01; 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.

Predictors	Total N	KOA case n (%)	Univariate OR (95% CI)	Multivariate OR (95% CI)
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)	1.04(0.86–1.25)
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) **	2.05 (1.74–2.41) **
Moderate-to-severe	165	50(30.3)	5.77(4.08–8.15) **	3.35(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) **	1.24(1.05–1.47) *
≥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

* $P \leq 0.05$, ** $P < 0.01$; 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.

Predictors	Total N	KOA case n (%)	Univariate OR (95% CI)	Multivariate OR (95% CI)
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	–
Low level	590	66(11.19)	1.17(0.89–1.54)	–
Middle-to-high level	2,661	269(10.11)	1.05(0.89–1.22)	–
MPA score				
No PA	3,192	278(8.71)	Reference	Reference
Low level	891	93(10.44)	1.22(0.95–1.56)	1.36(1.06–1.76) *
Middle-to-high level	4,110	444(10.8)	1.27(1.08–1.49) **	1.32 (1.12–1.56) **
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 level,	5,218	516(9.89)	1.01(0.84–1.22)	–
*P≤0.05, **P < 0.01; 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 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

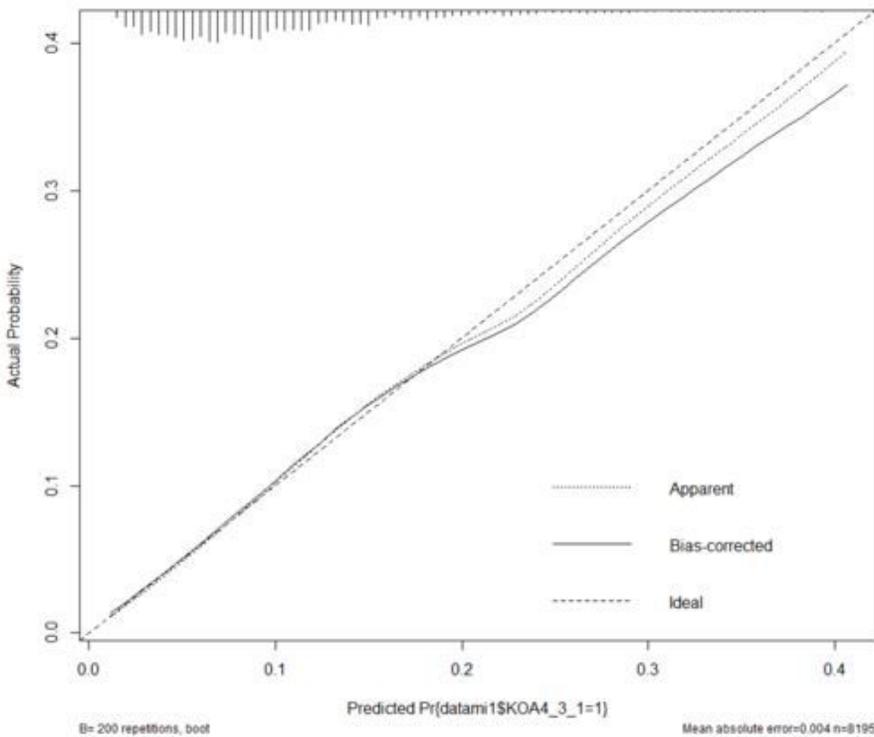
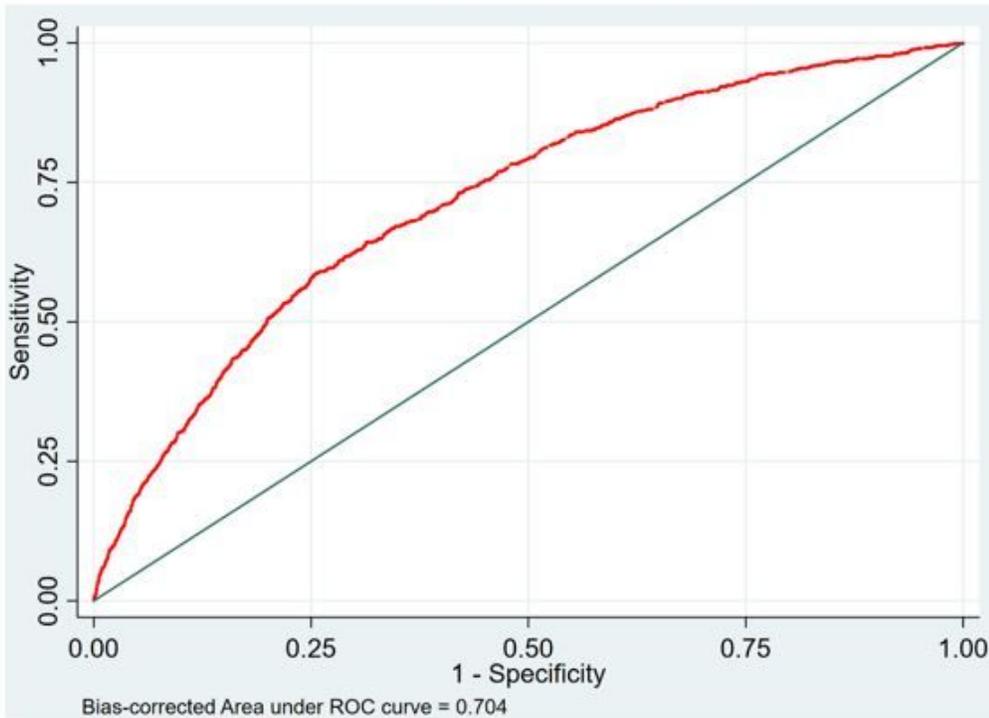


Figure 1

The discrimination and calibration curves of final risk model. (a): The discrimination curve 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. (b): The calibration curve. Area under the receiver characteristic curve, AUC.

Supplementary Files

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

- [SupplementaryBox1.docx](#)