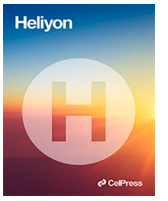
Heliyon 10 (2024) e24526

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Development and validation of a risk prediction model for arthritis in community-dwelling middle-aged and older adults in China

Mina Huang [a,b], Yue Guo [a], Zipeng Zhou [a], Chang Xu [a], Kun Liu [c], Yongzhu Wang [a], Zhanpeng Guo [a,*]

[a] Department of Orthopedics, The First Affiliated Hospital of Jinzhou Medical University, Jinzhou, China
[b] School of Nursing, Jinzhou Medical University, Jinzhou, China
[c] School of Medical College, Jinzhou Medical University, Jinzhou, China

ARTICLE INFO

Keywords:
Prediction model
Nomogram
Arthritis
Validation
Community-dwelling

ABSTRACT

Background: Considering its high prevalence, estimating the risk of arthritis in middle-aged and older Chinese adults is of particular interest. This study was conducted to develop a risk prediction model for arthritis in community-dwelling middle-aged and older adults in China.
Methods: Our study included a total of 9599 participants utilising data from the China Health and Retirement Longitudinal Study (CHARLS). Participants were randomly assigned to training and validation groups at a 7:3 ratio. Univariate and multivariate binary logistic regression analyses were used to identify the potential predictors of arthritis. Based on the results of the multivariate binary logistic regression, a nomogram was constructed, and its predictive performance was evaluated using the receiver operating characteristic (ROC) curve. The accuracy and discrimination ability were assessed using calibration curve analysis, while decision curve analysis (DCA) was performed to evaluate the net clinical benefit rate.
Results: A total of 9599 participants were included in the study, of which 6716 and 2883 were assigned to the training and validation groups, respectively. A nomogram was constructed to include age, hypertension, heart diseases, gender, sleep time, body mass index (BMI), residence address, the parts of joint pain, and trouble with body pains. The results of the ROC curve suggested that the prediction model had a moderate discrimination ability (AUC >0.7). The calibration curve of the prediction model demonstrated a good predictive accuracy. The DCA curves revealed a favourable net benefit for the prediction model.
Conclusions: The predictive model demonstrated good discrimination, calibration, and clinical validity, and can help community physicians and clinicians to preliminarily assess the risk of arthritis in middle-aged and older community-dwelling adults.

1. Introduction

Arthritis is a chronic inflammatory disease characterised by joint inflammation, synovial tissue swelling, joint stiffness, etc [1]. The most common types of arthritis are rheumatoid arthritis (RA), osteoarthritis (OA), psoriatic arthritis and inflammatory arthritis [2]. Arthritis has a high prevalence worldwide. An investigation demonstrated that the prevalence of arthritis in middle-aged and older

* Corresponding author.
E-mail address: ga83112@163.com (Z. Guo).

https://doi.org/10.1016/j.heliyon.2024.e24526
Received 30 August 2023; Received in revised form 5 January 2024; Accepted 10 January 2024
Available online 16 January 2024

adults in China was 31.4% [3]. In 2010–2012, 22.7% of all adults in the US were reported to have arthritis [4]. Studies have demonstrated a global prevalence of approximately 1% for RA [5,6]. Arthritis significantly affects the physical and mental well-being of individuals [7,8]. This could be reflected in physical pain [9,10], sleep limitations, high levels of stress and psychological distress [11]. Given the high incidence and significant impact of arthritis on an individual's quality of life, early detection is critical for the effective management of this chronic disease.

Arthritis was associated with a negative impact on the health of middle-aged and older adults [12]. Possible risk factors for arthritis included age and gender [3], obesity [13], perceived stress [14], and joint pain [15]. Researchers have pointed that population aging may be one of the predictors of arthritis [16]. The onset of the disease was most commonly observed in individuals aged 50–75 years [17]. Therefore, a risk-prediction model for arthritis in community-dwelling middle-aged and older adults may be useful for early detection and prevention. A model for predicting arthritis in community-dwelling individuals is therefore required.

Research on risk prediction models for arthritis has developed in the following aspects. Possible predictors in risk prediction models for knee osteoarthritis included age, female, BMI, occupational risks, family history, and knee injury [18]. The development of the stroke risk of prediction model for rheumatoid arthritis patients included possible predictors such as sex, age, systolic blood pressure, c-reactive protein [19]. The simplified version of the risk prediction model for knee osteoarthritis included possible predictors such as age, BMI, and knee injury [20]. Nevertheless, to the best of our knowledge, research on a risk-predictive model for arthritis in middle-aged and older adults in community-dwelling settings remains limited. Thus, it is necessary to develop an appropriate model that can facilitate the early identification of arthritis in community-dwelling middle-aged and older adults.

The development of arthritis risk prediction models that could allow early screening for arthritis and regulation of relevant factors, was an important measure for preventing arthritis in middle-aged and older adults in communities. Therefore, we aimed to develop a risk prediction model for arthritis in community-dwelling middle-aged and older adults.

2. Materials and methods

2.1. Study design and data source

Data were obtained from The China Health and Retirement Longitudinal Study (CHARLS), a nationally representative longitudinal survey in China. The CHARLS collects data on Chinese households and individuals aged ≥45 years to analyse the aging of the Chinese population. In 2011, 17,705 participants from 10,257 households were recruited across China's 28 provinces, including 150 counties or districts and 450 villages [21]. All participants were followed up every 2 years after the baseline survey. The data are publicly available (http://charls.pku.edu.cn/en), and we had no direct contact with the participants [22].

In this cross-sectional study, using data obtained from the CHARLS 2015 survey; data included information on demographic backgrounds, health status and functioning, health care, biomarkers, and blood-based bioassays [23]. The inclusion criteria for this study were: 1) individuals aged ≥45 years; 2) having data information regarding on having or not having arthritis. The exclusion criteria were: 1) persons aged <45 years; 2) missing values with more than 10% in the interested variables. Mean imputation was used to replace the missing values (<10%). A total of 9599 participants were included in this study. The participants were selected for screening in a randomized manner, maintaining a 7:3 ratio between the training and validation groups. Ultimately, 6716 (70%) and 2883 (30%) individuals were assigned to the training and validation groups, respectively. Participants were classified as having or not having arthritis.

2.2. Outcome and predictor variables

The outcome measure was the presence of arthritis. The answer to "Have you been diagnosed with arthritis or rheumatism by the doctor?" was defined as arthritis or no arthritis. Based on the answer to this question, we refer to a broad definition of arthritis [3], rather than a specific form of arthritis. This study used the demographic background, health status and functioning, blood-based bioassays, and biomarkers of participants as predictor variables. Fourteen predictor variables were considered in this study, including gender, age, marital status, trouble with body pains, the parts of joint pain, sleep time(h)/Day, health satisfaction, life satisfaction, hypertension, dyslipidaemia, diabetes, heart diseases, BMI, and residence address.

The demographic background factors included age, gender, residence address, and marital status. Gender was classified as either male or female. Residence address was defined as a city or town/village. Marital status was classified into three groups: married with spouse present, married but not living with spouse, and others (separated/divorced/widowed/never married/cohabitating).

Health status and functioning factors included trouble with body pains, the parts of joint pain, sleep time(h)/Day, health satisfaction, life satisfaction, hypertension, dyslipidaemia, diabetes, and heart diseases. Trouble with body pains was obtained from the question "Are you often troubled with any body pains?" which was answered by "yes" or "no". The parts of joint pain were obtained from the question "On what part of your body do you feel pain? Please list all parts of the body you are currently feeling pain." The parts of joint pain were classified as either "<2" or "≥2". Sleep time was obtained from the question "How many hours of actual sleep did you get at night?". Health satisfaction was obtained from the question "How satisfied are you with your health?" which was answered by "completely satisfied = 0, very satisfied = 1, somewhat satisfied = 2, not very satisfied = 3, not at all satisfied = 4". Life satisfaction was obtained from the question "Please think about your life as a whole. How satisfied are you with it?" which was answered by "completely satisfied = 0, very satisfied = 1, somewhat satisfied = 2, not very satisfied = 3, not at all satisfied = 4". Data on hypertension, dyslipidaemia, diabetes and heart diseases were obtained from the same question "Have you been diagnosed with [disesae] by the doctor?", which was answered by "yes" or "no". Biomarker factors included the BMI, obtained from the participant's

height and weight. BMI was classified as either "<30 kg/m^2" or "≥30 kg/m^2".

2.3. Statistical analysis

Statistical analyses were performed using IBM SPSS 25.0 and R4.2.2. First, continuous variables were presented as mean ± standard deviation, while categorical variables were reported as frequencies and percentages. Differences between groups were evaluated using the Chi-squared test or Fisher's exact test for categorical variables, and Student's t-test or Mann-Whitney *U* test for continuous variables. Univariate analysis was performed to screen for possible risk factors [24]. Second, collinearity was diagnosed for possible risk factors using variance inflation factors (VIF), with VIF <5 and tolerance >0.1 considered to indicate no significant collinearity [25]. Third, multivariate binary logistic regression analysis utilising forward stepwise regression was employed to identify the potential predictors of arthritis. The Hosmer-Lemeshow goodness-of-fit test was used to evaluate the degree of agreement of the prediction model, and $P > 0.05$ was identified as a good degree of predictive conformity [26]. The statistical significance level was $P < 0.05$.

The predictors obtained through multivariate binary logistic regression were selected to construct a nomogram using the "rms" and "regplot" packages in R software [27]. Receiver operating characteristic (ROC) curves were used to assess the predictive performance of the model. The evaluation indicators included area under the ROC curve (AUC), accuracy, sensitivity, and specificity [28]. An AUC ≥0.7 was considered satisfactory [29]. The calibration curve was generated through 1000 iterations of bootstrap self-sampling using the bootstrap method [30]. A calibration curve was constructed to assess the accuracy and discrimination ability of the models in the training and validation groups [31]. Decision curve analysis (DCA) was also performed using the "DecisionCurve" package in R software [27]. DCA was carried out to evaluate the net clinical benefit of the model [32]. The flowchart of the statistical analysis was shown in Fig. 1.

3. Results

3.1. Basic information

A total of 9599 participants (2900 with and 6699 without arthritis) were selected for this study (4546 [47.4%] males and 5053 [52.6%] females). The ages ranged from 45 to 94 years, and the mean age was 59.95 ± 9.46 years. As shown in Table 1, the participants were randomly divided into training and validation groups at a ratio of 7:3. There were no differences in baseline characteristics between the training group and the validation group except for residence address.

A total of 6716 participants (arthritis: 2045, no arthritis: 4671) were enrolled in the training group, of whom 3207 (47.8%) and 3509 (52.2%) were males and females, respectively. The number of females with arthritis was higher than males, and the mean age was 59.86 ± 9.44 years. The prevalence of arthritis was 30.4%. The validation group comprised 2883 participants (arthritis: 855, no arthritis: 2028), including 1339 (46.4%) males and 1544 (53.6%) females, with an average age of 60.15 ± 9.51 years. The prevalence of arthritis was 29.7%.

3.2. Univariate analysis and multicollinearity diagnosis

Table 2 presented the basic characteristics of the participants in the training group. Univariate analysis (Table 2) showed that gender, age, marital status, trouble with body pains, the parts of joint pain, sleep time(h)/Day, health satisfaction, life satisfaction, hypertension, dyslipidemia, diabetes, heart diseases, BMI and residence address were correlated with arthritis ($P < 0.05$). These factors may be potential risk factors for arthritis in middle-aged and older community-dwelling adults.

The variables that exhibited a significance level of $P < 0.05$ in the univariate analysis were assessed for multicollinearity to ascertain if there was multicollinearity among them. Table 3 showed the collinearity of the diagnostic results for these variables. The results showed that the VIF of the variables was <5 and the tolerance was >0.1, indicating that there was no multicollinearity among the variables.

3.3. Multivariate binary logistic regression analysis

The 14 variables that passed the screening in the univariate analysis and multicollinearity diagnosis were selected for multivariate

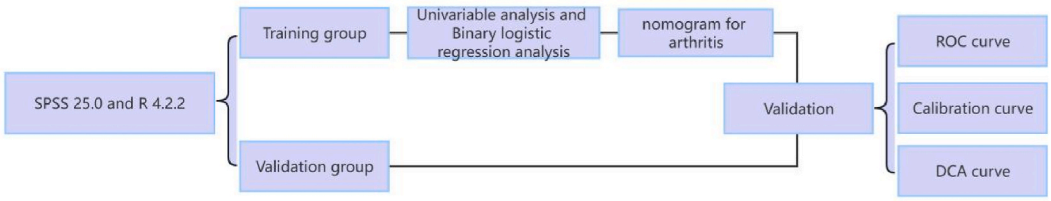


Fig. 1. Flow diagram of analysis.

Table 1
Characteristics of baseline demographic and clinical indicators of the participants in the different groups.

Variables	Training group N(6716)	Validation group N(2883)	χ^2/t	*P*-value
	N (%)/Mean(SD)			
Gender			1.382	0.240
Male	3207(47.8)	1339(46.4)		
Female	3509(52.2)	1544(53.6)		
Age(years)	59.86(9.44)	60.15(9.51)	−0.421	0.674
Marital status			1.293	0.524
Married with Spouse Present	5586(83.2)	2413(83.7)		
Married But Not Living with Spouse	295(4.4)	112(3.9)		
others	835(12.4)	358(12.4)		
Trouble with body pains			0.001	0.981
No	5499(81.9)	2360(81.9)		
Yes	1217(18.1)	523(18.1)		
The parts of joint pain			0.177	0.674
<2	5815(86.6)	2487(86.3)		
≥2	901(13.4)	396(13.7)		
Sleep time(h)/Day	6.49(1.84)	6.46(1.83)	0.819	0.413
Health satisfaction			5.258	0.262
Completely Satisfied	254(3.8)	85(2.9)		
Very Satisfied	1417(21.1)	593(20.6)		
Somewhat Satisfied	3416(50.9)	1505(52.2)		
Not Very Satisfied	1276(19.0)	541(18.8)		
Not at All Satisfied	353(5.3)	159(5.5)		
Life satisfaction			0.234	0.994
Completely Satisfied	427(6.4)	183(6.3)		
Very Satisfied	2446(36.4)	1046(36.3)		
Somewhat Satisfied	3330(49.6)	1430(49.6)		
Not Very Satisfied	417(6.2)	185(6.4)		
Not at All Satisfied	96(1.4)	39(1.4)		
Hypertension			1.463	0.226
No	5232(77.9)	2278(79.0)		
Yes	1484(22.1)	605(21.0)		
Dyslipidemia			0.130	0.718
No	6050(90.1)	2604(90.3)		
Yes	666(9.9)	279(9.7)		
Diabetes			0.006	0.939
No	6343(94.4)	2724(94.5)		
Yes	373(5.6)	159(5.5)		
Heart diseases			2.420	0.120
No	5939(88.4)	2581(89.5)		
Yes	777(11.6)	302(10.5)		
BMI(kg/m^2)			3.056	0.080
<30	6372(94.9)	2710(94.0)		
≥30	344(5.1)	173(6.0)		
Residence Address			6.041	0.014[a]
City	1140(17.0)	431(14.9)		
Town/village	5576(83.0)	2452(85.1)		
Arthritis			0.602	0.438
No	4671(69.6)	2028(70.3)		
Yes	2045(30.4)	855(29.7)		

Note: *P*-values from Chi-squared tests or t-tests. Chi-squared tests: 0 cells (0.0%) have an expected count less than 5.
Others (separated or divorced or widowed or never married or cohabitated).
[a] Indicates statistical significance.

binary logistic regression analysis using forward stepwise regression to identify the predictive factors for arthritis. With the presence or absence of arthritis as the dependent variable, no arthritis = 0 and arthritis = 1. Fourteen variables as the independent variables. Multivariate binary logistic regression analysis was used to further explore the influence of independent variables on the presence or absence of arthritis. As shown in Table 4, nine variables were entered into the logistic regression equation: age, hypertension, heart diseases, gender, sleep time(h)/Day, BMI, residence address, the parts of joint pain and trouble with body pains. The result of the Hosmer-Lemeshow test indicated that the prediction model had a good degree of fit ($\chi^2 = 9.567$, $P = 0.297$). The results of multivariate logistic regression analysis showed that: age, hypertension, heart diseases, gender, sleep time(h)/Day, BMI, residence address, the parts of joint pain, and trouble with body pains were independent predictive factors for arthritis (Table 4).

Table 2
Univariate analysis of baseline demographic and clinical indicators of the participants in the training group.

Variables	Arthritis N(2045)	No arthritis N(4671)	χ^2/t	*P*-value
	N (%)/Mean(SD)			
Gender			52.523	<0.001
Male	840(41.1)	2367(50.7)		
Female	1205(58.9)	2304(49.3)		
Age(years)	61.86(8.85)	58.99(9.56)	−11.943	<0.001
Marital status			21.534	<0.001
Married with Spouse Present	1646(80.5)	3940(84.4)		
Married But Not Living with Spouse	87(4.3)	208(4.5)		
others	312(15.3)	523(11.2)		
Trouble with body pains			576.612	<0.001
No	1063(52.0)	3765(80.6)		
Yes	982(48.0)	906(19.4)		
The parts of joint pain			729.749	<0.001
<2	1282(62.7)	4217(90.3)		
≥2	763(37.3)	454(9.7)		
Sleep time(h)/Day	6.15(2.00)	6.64(1.75)	9.735	<0.001
Health satisfaction			202.133	<0.001
Completely Satisfied	53(2.6)	201(4.3)		
Very Satisfied	306(15.0)	1111(23.8)		
Somewhat Satisfied	982(48.0)	2434(52.1)		
Not Very Satisfied	529(25.9)	747(16.0)		
Not at All Satisfied	175(8.6)	178(3.8)		
Life satisfaction			98.184	<0.001
Completely Satisfied	96(4.7)	331(7.1)		
Very Satisfied	623(30.5)	1823(39.0)		
Somewhat Satisfied	1104(54.0)	2226(47.7)		
Not Very Satisfied	171(8.4)	246(5.3)		
Not at All Satisfied	51(2.5)	45(1.0)		
Hypertension			98.288	<0.001
No	1438(70.3)	3794(81.2)		
Yes	607(29.7)	877(18.8)		
Dyslipidemia			15.380	<0.001
No	1798(87.9)	4252(91.0)		
Yes	247(12.1)	419(9.0)		
Diabetes			24.123	<0.001
No	1889(92.4)	4454(95.4)		
Yes	156(7.6)	217(4.6)		
Heart diseases			151.355	<0.001
No	1660(81.2)	4279(91.6)		
Yes	385(18.8)	392(8.4)		
BMI(kg/m^2)			4.307	0.038
<30	1923(94.0)	4449(95.2)		
≥30	122(6.0)	222(4.8)		
Residence Address			44.200	<0.001
City	253(12.4)	887(19.0)		
Town/village	1792(87.6)	3784(81.0)		

Note: *P*-values from Chi-squared tests or t-tests. Chi-squared tests: 0 cells (0.0%) have an expected count less than 5.
Others (separated or divorced or widowed or never married or cohabitated).

3.4. Construction and validation of a nomogram for arthritis

Based on the results of multivariate binary logistic regression analysis, a nomogram was constructed using nine independent predictive variables for arthritis, as shown in Fig. 2. Fig. 2 showed that the total score of the nomogram was between 0 and 400, and the probability of arthritis risk was between 0.1 and 0.9. The risk of arthritis was calculated by adding the corresponding scores in the nomogram for age, hypertension, heart diseases, gender, sleep time, BMI, residence address, the parts of joint pain and trouble with body pains, summed to count the total score. The higher the total score, the greater the risk of arthritis in middle-aged and older adults in the community. By applied this nomogram to a 64-year-old village man who had no hypertension, no heart diseases, no trouble with joint pains, BMI <30 kg/m^2, sleep time was 7 h/day, and the parts of joint pain <2. We can calculate that his total score is 325, which corresponds to a 19.2% risk of developing arthritis (Fig. 2).

After constructing the nomogram prediction model, the ROC curves were used to estimate the discrimination of the nomogram. The ROC curves for the training and validation groups were shown in Fig. 3. The results for the training group in Fig. 3A and Table 5 showed that the AUC, the Youden index, sensitivity, and specificity were 0.723 (95%CI: 0.710–0.737), 0.326, 0.565, and 0.761, respectively. The results for the validation group in Fig. 3B and Table 5 showed that the AUC, the Youden index, sensitivity, and specificity were 0.721 (95%CI: 0.700–0.741), 0.323, 0.556, and 0.767, respectively. Therefore, the results of both groups suggested

Table 3
Multicollinearity diagnosis of predictors.

Factors	Tolerance	VIF
Gender	0.947	1.056
Age(years)	0.869	1.150
Marital status	0.913	1.095
Trouble with body pains	0.419	2.387
The parts of joint pain	0.430	2.327
Sleep time(h)/Day	0.948	1.054
Health satisfaction	0.780	1.283
Life satisfaction	0.838	1.193
Hypertension	0.832	1.201
Dyslipidemia	0.861	1.161
Diabetes	0.893	1.119
Heart diseases	0.899	1.112
BMI	0.965	1.036
Residence Address	0.974	1.027

Table 4
The results of the binary logistic regression analysis in the training group.

Variables	B	SE	Wald	*P*	OR	95%CI
Age(years)	0.030	0.003	94.242	<0.001	1.031	1.024–1.037
Hypertension	0.300	0.069	18.667	<0.001	1.349	1.178–1.546
Heart diseases	0.567	0.087	42.637	<0.001	1.763	1.487–2.090
Gender	0.214	0.059	13.181	<0.001	1.239	1.104–1.390
Sleep time(h)/Day	−0.079	0.016	25.600	<0.001	0.924	0.896–0.953
BMI	0.267	0.128	4.334	0.037	1.306	1.016–1.679
Residence Address	0.450	0.083	29.712	<0.001	1.569	1.334–1.844
The parts of joint pain	1.173	0.104	126.889	<0.001	3.230	2.634–3.961
Trouble with body pains	0.446	0.092	23.449	<0.001	1.563	1.304–1.872
constants	−3.194	0.235	184.172	<0.001	0.041	

Notes: OR: Odds Ratio. 95%CI: 95% Confidence Interval.

that the nomogram prediction model had moderate discrimination ability (AUC >0.7). Also, Table 5 showed that the performance of the nomogram in the training and validation groups were superior to the single factor.

The calibration curve of the nomogram prediction model for arthritis was shown in Fig. 4(A and B). The abscissa represented the probability of arthritis predicted by the nomogram, and the ordinate represented the actual probability of arthritis. The ideal line indicated that the predicted probability was equal to the actual probability under optimal conditions. The apparent and bias-corrected lines represented the predicted and actual probabilities of the nomogram prediction model under realistic conditions, respectively. The closer the apparent and bias-corrected lines were to the ideal line, the better the calibration of the nomogram prediction model, and the better the prediction model. As shown in Fig. 4(A and B), the prediction accuracy of this prediction model was good.

Fig. 5(A and B) showed the DCA curves for the training and validation groups. The results demonstrated that the nomogram risk-prediction model provided clinical benefits to participants across thresholds ranging from 0 to 0.8. The analysis indicated that the prediction model yielded a good net benefit.

4. Discussion

In this study, we developed and validated a prediction model for arthritis that could be used to assess the risk of arthritis in community-dwelling middle-aged and older adults. The model incorporated information which included predictors such as age, hypertension, heart diseases, gender, sleep time(h)/Day, BMI, residence address, the parts of joint pain and trouble with body pains.

The results of this study suggested that the parts of joint pain was considered as the most important predictor for arthritis. In this regard, it was known that the joint pain indicated a risk of arthritis [33]. Meanwhile, joint pain could also be used as one of the predictors to differentiate inflammatory arthritis and noninflammatory arthritis [34]. With the acceleration of aging, joint pain caused by arthritis, especially osteoarthritis may lead to difficulties in middle-aged and elderly individuals [35]. Trouble with body pains was also one of the predictors in the model. One study revealed that arthritis was correlated with body pain and was one of the predictors of older adults reporting body pain [36]. This finding was consistent with the results of the present study. Thus, as a risk predictor of arthritis, pain should deserve the attention of community medical workers, as well as middle-aged and older adults.

As age increases, the body and tissue functions gradually decrease. Various studies have demonstrated a correlation between age and arthritis [4,37,38]. The results of the present study showed that the risk of developing arthritis increased with age in middle-aged and older adults in communities. This was consistent with findings regarding the effect of age proposed in the risk prediction model for knee osteoarthritis [20,39]. These findings indicated that age was one of the factors in the risk prediction model for arthritis. Gender was also observed as a risk factor of arthritis. This study identified gender as one of the model predictors, with females demonstrating a

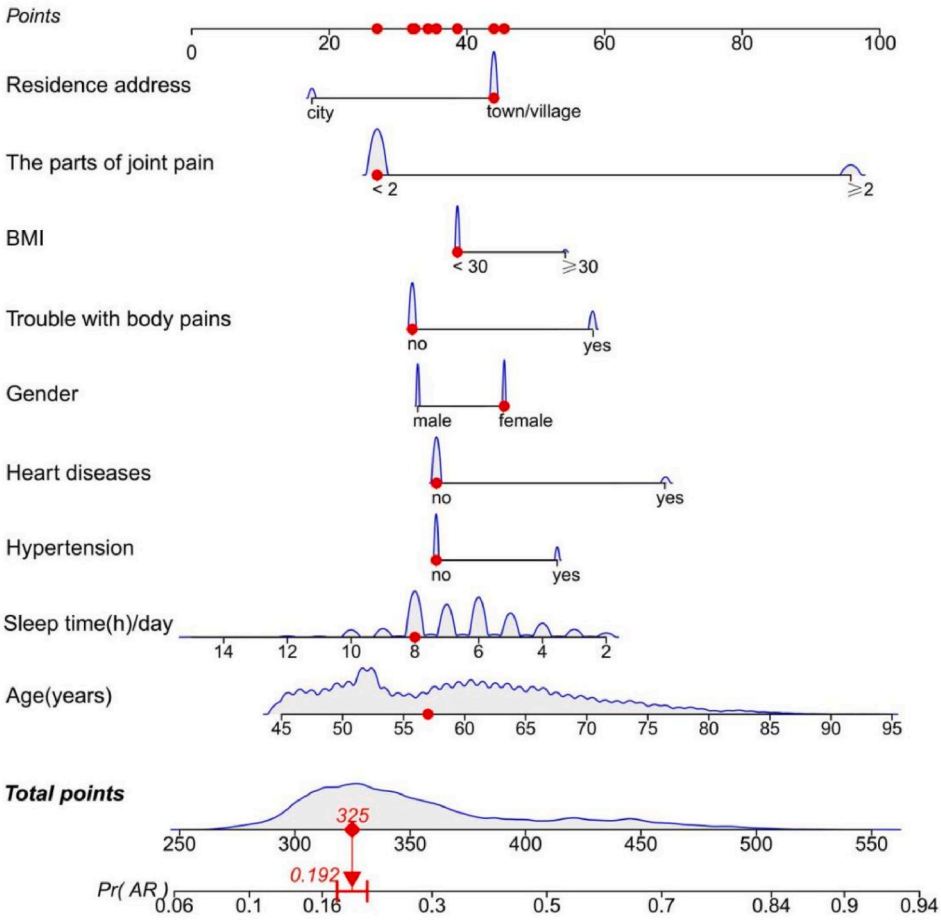


Fig. 2. Nomogram for a risk prediction of arthritis in the training group.

higher risk of arthritis than males. Studies have also shown that female was associated with having osteoarthritis [40–42]. These results were consistent with those of the present study. Additionally, we found that residence address was also a risk predictor for arthritis. People in rural areas were found to have a higher risk of arthritis than middle-aged and older adults in cities. Another study also reported that residential area (urban or rural) was a predictive factor for the risk of symptomatic knee osteoarthritis [39], which is consistent with the results of the present study. This might be related to the environmental risk factors for arthritis [43]. Thus, older females in rural areas might be at higher risk of arthritis and require more attention from community healthcare workers.

Heart diseases mainly including heart attack, coronary heart disease, angina, heart failure, and other heart problems were also another predictive factor for arthritis. Multivariate analysis of the study showed that the risk of arthritis in middle-aged and older adults with heart diseases was 0.763 times higher than those without heart diseases. Research indicated that cardiac manifestations could increase the incidence rate of inflammatory joint diseases [44]. Patients with inflammatory joint diseases had an increased risk of cardiovascular disease [45]. Hypertension was also shown to be a risk predictor of arthritis. Studies indicated that arthritis was strongly associated with hypertension [46,47]. Therefore, heart diseases and hypertension in communities had a significant effect on the risk of arthritis.

The present study also confirmed that obesity (BMI $\geq$30 kg/m^2) was a risk factor for arthritis. Obesity was associated with a higher risk of arthritis compared with a BMI <30 kg/m^2, which was consistent with the findings of previous studies demonstrating that obesity increases the risk of arthritis [48,49]. Interestingly, sleep time was confirmed to be a negative predictor of arthritis. The risk of arthritis decreased with longer sleep time. Prior research indicated that a short sleep duration was associated with an increased risk of rheumatoid arthritis [50]. This was also reflected in our study. Therefore, middle-aged and older adults could increase their sleep time appropriately.

Through a rigorous literature search, there was no relevant literature to develop a nomogram based on the risk of arthritis in community-dwelling middle-aged and older adults. We identified several predictors that were associated with the risk of arthritis in community-dwelling middle-aged and older adults. We found that the combination index outperformed a single marker in terms of predictive ability [51]. Therefore, we developed a nomogram based on these variables to predict the risk of arthritis in community-dwelling middle-aged and older adults. The nomogram that was constructed exhibited exceptional discrimination and

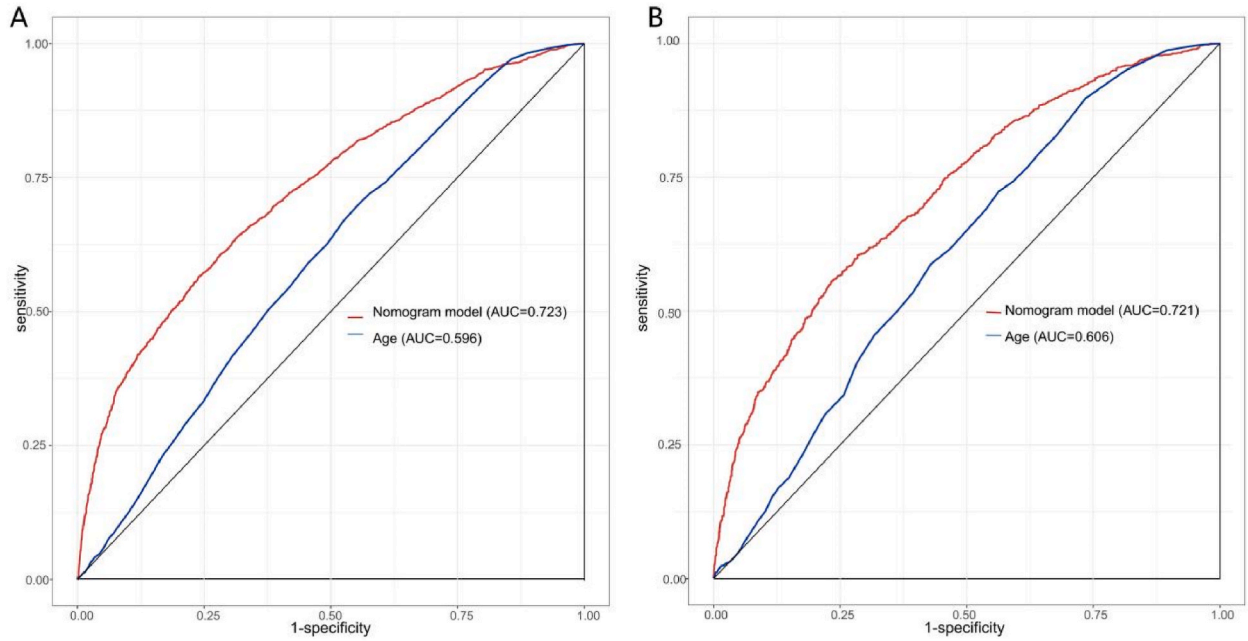


Fig. 3. Discrimination of the nomogram prediction model for arthritis. (A) ROC curve of the training group. (B) ROC curve of the validation group.

Table 5
The performance of the nomogram in the training and validation groups.

	AUC (95% CI)	Youden index	Sensitivity	Specificity
Training group				
Age(years)	0.596(0.581–0.610)	0.145	0.7	0.445
Nomogram	0.723(0.710–0.737)	0.326	0.565	0.761
Validation group				
Age(years)	0.606(0.585–0.627)	0.163	0.897	0.266
Nomogram	0.721 (0.700–0.741)	0.323	0.556	0.767

AUC, area under the curve; CI, confidence interval.

calibration abilities in both the training and validation groups. Then decision curve analysis was employed to assess the net clinical benefits of the nomogram for guiding clinical decisions. The results revealed that the net clinical benefits provided by the nomogram surpassed those offered by a solitary marker in both the training and validation groups. The analysis indicated that the nomogram yielded a good net benefit. These results demonstrate that the nomogram can accurately predict the risk of arthritis in community-dwelling middle-aged and older adults.

The above discussion regarding the predictors of the risk of arthritis had its potential clinical implications. It has several strengths. First, the data were obtained from a nationwide representative sample. Second, the prediction model for arthritis in middle-aged and older adults in the community has certain effects on health guidance and prevention. Still, this study has several limitations. First, although the data from middle-aged and older adults in the community were representative, external validation of the clinical data is still lacking. So clinical data should be used for external validation in the future. Second, the variables were derived from self-reports, which might lead to bias. Nevertheless, the large sample size, well-designed questionnaire, and random selection of training and validation groups may have reduced bias to a certain extent. Third, the survey did not consider longitudinal data, which may have affected the validity of the prediction model to some extent. Therefore, follow-up studies that use longitudinal surveys are warranted. Further studies are needed to analyse predictive models for different types of arthritis.

5. Conclusions

In conclusion, we developed and validated a risk prediction model for arthritis in community-dwelling middle-aged and older adults. The prediction model had a high sensitivity and specificity, as well as good discrimination and calibration abilities. The predictive model will facilitate community health institutions and clinicians in screening and predicting the incidence of arthritis among community residents, while also enabling the implementation of early intervention measures.

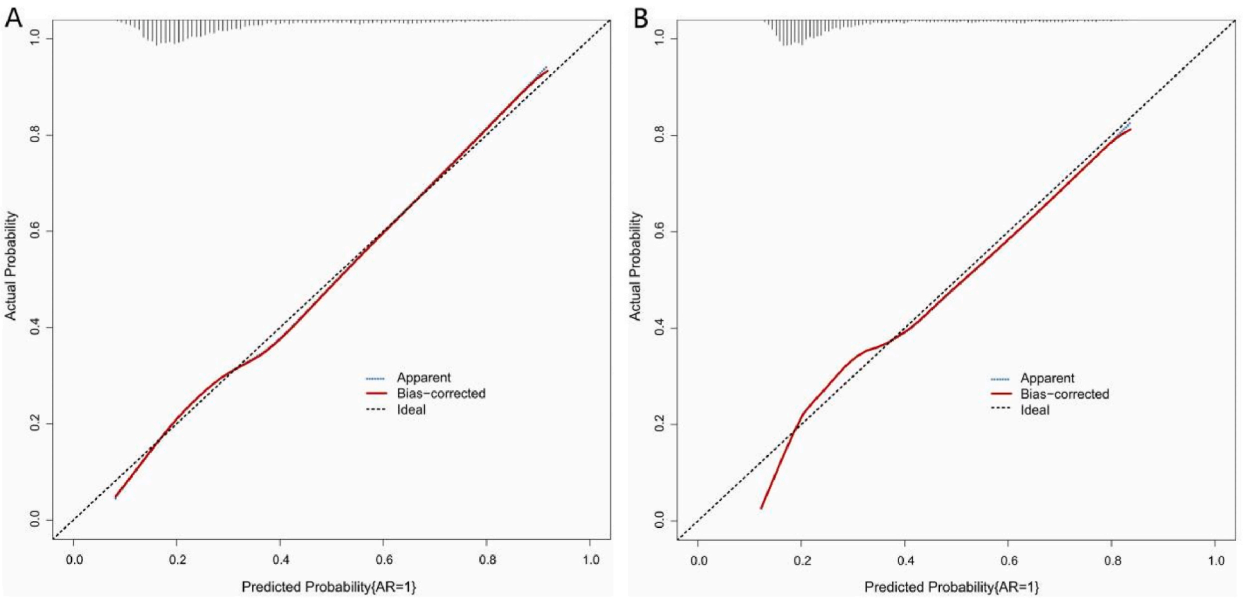


Fig. 4. Calibration curves of the nomogram prediction models for arthritis. (A) Calibration curve of the training group. (B) Calibration curve of the validation group.

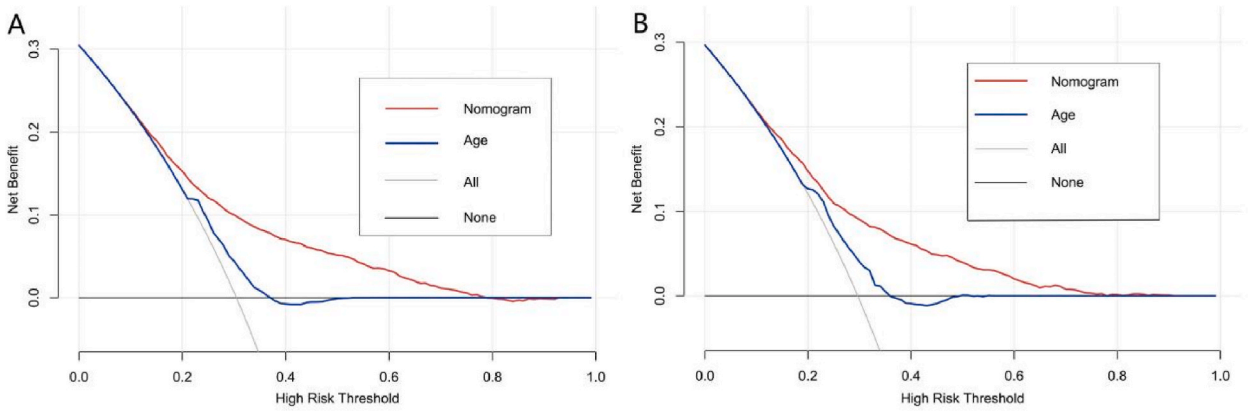


Fig. 5. DCA of the nomogram prediction models for arthritis. (A) DCA of the training group. (B) DCA of the validation group. The horizontal axis represented the risk threshold probability. The vertical axis represented the net benefit. DCA: Decision Curve Analysis.

Data availability statement

The data presented in this study are openly available from the China Health and Retirement Longitudinal Study. The data are available at: https://charls.charlsdata.com/pages/Data/2015-charls-wave4/zh-cn.html

Funding

No financial support was received for this study.

Additional information

No additional information is available for this paper.

CRediT authorship contribution statement

Mina Huang: Writing – original draft, Formal analysis, Data curation. **Yue Guo:** Writing – review & editing. **Zipeng Zhou:** Writing

– review & editing. **Chang Xu:** Writing – review & editing. **Kun Liu:** Supervision, Methodology. **Yongzhu Wang:** Formal analysis. **Zhanpeng Guo:** Writing – review & editing, Supervision, Software, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We are grateful to all participants enrolled in the CHARLS and its members. We are grateful to CHARLS for providing us with these data. We appreciate all the co-researchers, reviewers, and editors.

References

[1] J.J. Goronzy, C.M. Weyand, Developments in the scientific understanding of rheumatoid arthritis, Arthritis Res. Ther. 11 (5) (2009) 249, https://doi.org/10.1186/ar2758.
[2] C.H. Tang, Research of pathogenesis and novel therapeutics in arthritis, Int. J. Mol. Sci. 20 (7) (2019), https://doi.org/10.3390/ijms20071646.
[3] C. Li, T. Liu, W. Sun, L. Wu, Z.Y. Zou, Prevalence and risk factors of arthritis in a middle-aged and older Chinese population: the China health and retirement longitudinal study, Rheumatology 54 (4) (2015) 697–706, https://doi.org/10.1093/rheumatology/keu391.
[4] J.M. Hootman, C.G. Helmick, K.E. Barbour, K.A. Theis, M.A. Boring, Updated projected prevalence of self-reported doctor-diagnosed arthritis and arthritis-attributable activity limitation among US adults, 2015-2040, Arthritis Rheumatol. 68 (7) (2016) 1582–1587, https://doi.org/10.1002/art.39692.
[5] A. Wasserman, Rheumatoid arthritis: common questions about diagnosis and management, Am. Fam. Physician 97 (7) (2018) 455–462.
[6] P. Prasad, S. Verma, Surbhi, N.K. Ganguly, V. Chaturvedi, S.A. Mittal, Rheumatoid arthritis: advances in treatment strategies, Mol. Cell. Biochem. 478 (1) (2023) 69–88, https://doi.org/10.1007/s11010-022-04492-3.
[7] B. Stubbs, et al., Lifetime self-reported arthritis is associated with elevated levels of mental health burden: a multi-national cross sectional study across 46 low- and middle-income countries, Sci. Rep. 7 (1) (2017) 7138, https://doi.org/10.1038/s41598-017-07688-6.
[8] W. Kang, Global and dimensions of mental health in arthritis patients, Healthcare (Basel, Switzerland) 11 (2) (2023), https://doi.org/10.3390/healthcare11020195.
[9] M. Harth, W.R. Nielson, Pain and affective distress in arthritis: relationship to immunity and inflammation, Expet Rev. Clin. Immunol. 15 (5) (2019) 541–552, https://doi.org/10.1080/1744666X.2019.1573675.
[10] A.L. Horgas, A.L. Elliott, S. Yang, Y. Guo, Cross-sectional relationship between pain intensity and subjective cognitive decline among middle-aged and older adults with arthritis or joint conditions: results from a population-based study, SAGE Open Med. 10 (2022) 20503121221095923, https://doi.org/10.1177/20503121221095923.
[11] S. O'Donnell, C. Rusu, G.A. Hawker, et al., Arthritis has an impact on the daily lives of Canadians young and old: results from a population-based survey, BMC Muscoskel. Disord. 16 (2015) 230, https://doi.org/10.1186/s12891-015-0691-2.
[12] R. Yang, J. Wang, H. Wang, E.L. Tracy, C.T. Tracy, A cross-lagged model of depressive symptoms and mobility disability among middle-aged and older Chinese adults with arthritis, Geriatr. Gerontol. Int. 20 (10) (2020) 873–877, https://doi.org/10.1111/ggi.13993.
[13] K.R. Fontaine, S. Haaz, S.J. Bartlett, Are overweight and obese adults with arthritis being advised to lose weight, J. Clin. Rheumatol. 13 (1) (2007) 12–15, https://doi.org/10.1097/01.rhu.0000256168.74277.15.
[14] M.L. Harris, D. Loxton, D.W. Sibbritt, J.E. Byles, The influence of perceived stress on the onset of arthritis in women: findings from the Australian Longitudinal Study on women's health, Ann. Behav. Med. 46 (1) (2013) 9–18, https://doi.org/10.1007/s12160-013-9478-6.
[15] D. Guglielmo, et al., State-specific severe joint pain and physical inactivity among adults with arthritis - United States, 2017, MMWR. Morb. Mortal. Weekly Rep. 68 (17) (2019) 381–387, https://doi.org/10.15585/mmwr.mm6817a2.
[16] J.M. Hootman, C.G. Helmick, Projections of US prevalence of arthritis and associated activity limitations, Arthritis Rheum. 54 (1) (2006) 226–229, https://doi.org/10.1002/art.21562.
[17] L. Lufkin, M. Budišić, S. Mondal, S. Sur, A bayesian model to analyze the association of rheumatoid arthritis with risk factors and their interactions, Front. Public Health 9 (2021) 693830, https://doi.org/10.3389/fpubh.2021.693830.
[18] W. Zhang, et al., Nottingham knee osteoarthritis risk prediction models, Ann. Rheum. Dis. 70 (9) (2011) 1599–1604, https://doi.org/10.1136/ard.2011.149807.
[19] F. Xin, et al., Development and validation of a nomogram for predicting stroke risk in rheumatoid arthritis patients, Aging (Albany NY) 13 (11) (2021) 15061–15077, https://doi.org/10.18632/aging.203071.
[20] K. Magnusson, A. Turkiewicz, S. Timpka, M. Englund, A prediction model for the 40-year risk of knee osteoarthritis in adolescent men, Arthritis Care Res. 71 (4) (2019) 558–562, https://doi.org/10.1002/acr.23685.
[21] Y. Zhao, Y. Hu, J.P. Smith, J. Strauss, G. Yang, Cohort profile: the China health and retirement longitudinal study (CHARLS), Int. J. Epidemiol. 43 (1) (2014) 61–68, https://doi.org/10.1093/ije/dys203.
[22] L. Zhang, et al., Combined effect of famine exposure and obesity parameters on hypertension in the midaged and older adult: a population-based cross-sectional study, BioMed Res. Int. 2021 (2021) 5594718, https://doi.org/10.1155/2021/5594718.
[23] Y. Zhang, et al., The prevalence of obesity-related hypertension among middle-aged and older adults in China, Front. Public Health 10 (2022) 865870, https://doi.org/10.3389/fpubh.2022.865870.
[24] D. Xue, et al., Risk factor analysis and a predictive model of postoperative depressive symptoms in elderly patients undergoing video-assisted thoracoscopic surgery, Brain Sci. 13 (4) (2023), https://doi.org/10.3390/brainsci13040646.
[25] Q.M. Jiang, et al., Predictors and dynamic nomogram to determine the individual risk of malignant brain edema after endovascular thrombectomy in acute ischemic stroke, J. Clin. Neurol. 18 (3) (2022) 298–307, https://doi.org/10.3988/jcn.2022.18.3.298.
[26] X. Yan, et al., Development and assessment of a risk prediction model for moderate-to-severe obstructive sleep apnea, Front. Neurosci. 16 (2022) 936946, https://doi.org/10.3389/fnins.2022.936946.
[27] A. Jiang, et al., Establishment and validation of a nomogram to predict the in-hospital death risk of nosocomial infections in cancer patients, Antimicrob. Resist. Infect. Control 11 (1) (2022) 29, https://doi.org/10.1186/s13756-022-01073-3.
[28] S.F. Alhabib, I. Saliba, Reliability of monothermal caloric test as screening test of vestibular system, J. Clin. Med. 11 (23) (2022), https://doi.org/10.3390/jcm11236977.
[29] C. Romero-Blanco, A. Hernández-Martínez, M.L. Parra-Fernández, M.D. Onieva-Zafra, M. Prado-Laguna, J. Rodríguez-Almagro, Food addiction and lifestyle habits among university students, Nutrients 13 (4) (2021), https://doi.org/10.3390/nu13041352.
[30] X. Dai, M. Yuan, M. Dang, D. Liu, W. Fei, Development and validation of a predictive model for chronic postsurgical pain after arthroscopic rotator cuff repair: a prospective cohort study, J. Pain Res. 16 (2023) 3273–3288, https://doi.org/10.2147/JPR.S423110.

[31] Y. Zhou, et al., Identification of copper death-associated molecular clusters and immunological profiles in rheumatoid arthritis, Front. Immunol. 14 (2023) 1103509, https://doi.org/10.3389/fimmu.2023.1103509.
[32] J. Lin, et al., Development and validation of survival nomograms in patients with primary bladder lymphoma, J. Clin. Med. 11 (11) (2022), https://doi.org/10.3390/jcm11113188.
[33] J.J. McDougall, Arthritis and pain. Neurogenic origin of joint pain, Arthritis Res. Ther. 8 (6) (2006) 220, https://doi.org/10.1186/ar2069.
[34] M.C. Lu, M. Koo, N.S. Lai, Clinimetric properties of the Chinese version of the early inflammatory arthritis detection tool, BMC Muscoskel. Disord. 16 (2015) 243, https://doi.org/10.1186/s12891-015-0706-z.
[35] X. Sun, et al., Osteoarthritis in the middle-aged and elderly in China: prevalence and influencing factors, Int. J. Environ. Res. Publ. Health 16 (23) (2019), https://doi.org/10.3390/ijerph16234701.
[36] L. Yang, W. Peng, Prevalence and factors associated with body pain: results of a nationally representative survey of 9,586 Chinese adults aged 60 and over, Front. Public Health 9 (2021) 634123, https://doi.org/10.3389/fpubh.2021.634123.
[37] M.B. Hossain, J.A. Kopec, M. Atiquzzaman, M.E. Karim, The association between rheumatoid arthritis and cardiovascular disease among adults in the United States during 1999-2018, and age-related effect modification in relative and absolute scales, Ann. Epidemiol. 71 (2022) 23–30, https://doi.org/10.1016/j.annepidem.2022.03.005.
[38] R.C. Stone, J. Baker, Physical activity, age, and arthritis: exploring the relationships of major risk factors on biopsychosocial symptomology and disease status, J. Aging Phys. Activ 22 (3) (2014) 314–323, https://doi.org/10.1123/japa.2012-0293.
[39] L. Wang, H. Lu, H. Chen, S. Jin, M. Wang, S. Shang, Development of a model for predicting the 4-year risk of symptomatic knee osteoarthritis in China: a longitudinal cohort study, Arthritis Res. Ther. 23 (1) (2021) 65, https://doi.org/10.1186/s13075-021-02447-5.
[40] C.M. Bertoncelli, et al., Predicting osteoarthritis in adults using statistical data mining and machine learning, Ther. Adv. Musculoskelet. Dis. 14 (2022), https://doi.org/10.1177/1759720X221104935, 1759720X221104935.
[41] E. Grossi, Do artificial neural networks love sex? How the combination of artificial neural networks with evolutionary algorithms may help to identify gender influence in rheumatic diseases, Clin. Exp. Rheumatol. 41 (1) (2023) 1–5, https://doi.org/10.55563/clinexprheumatol/vgl2nz.
[42] E.G. Favalli, M. Biggioggero, C. Crotti, A. Becciolini, M.G. Raimondo, P.L. Meroni, Sex and management of rheumatoid arthritis, Clin. Rev. Allergy Immunol. 56 (3) (2019) 333–345, https://doi.org/10.1007/s12016-018-8672-5.
[43] P. Iltchev, et al., Epidemiology of Rheumatoid Arthritis (RA) in rural and urban areas of Poland - 2008-2012, Ann. Agric. Environ. Med. 23 (2) (2016) 350–356, https://doi.org/10.5604/12321966.1203904.
[44] S. Castaneda, C. Gonzalez-Juanatey, M.A. Gonzalez-Gay, Inflammatory arthritis and heart disease, Curr. Pharmaceut. Des. 24 (3) (2018) 262–280, https://doi.org/10.2174/1381612824666180123102632.
[45] G. Wibetoe, et al., Cardiovascular disease risk profiles in inflammatory joint disease entities, Arthritis Res. Ther. 19 (1) (2017) 153, https://doi.org/10.1186/s13075-017-1358-1.
[46] X. Liang, O. Chou, C.L. Cheung, B. Cheung, Is hypertension associated with arthritis? The United States national health and nutrition examination survey 1999-2018, Ann. Med. 54 (1) (2022) 1767–1775, https://doi.org/10.1080/07853890.2022.2089911.
[47] B. Hadwen, S. Stranges, L. Barra, Risk factors for hypertension in rheumatoid arthritis patients-A systematic review, Autoimmun. Rev. 20 (4) (2021) 102786, https://doi.org/10.1016/j.autrev.2021.102786.
[48] D. Ye, Y. Mao, Y. Xu, X. Xu, Z. Xie, C. Wen, Lifestyle factors associated with incidence of rheumatoid arthritis in US adults: analysis of National Health and Nutrition Examination Survey database and meta-analysis, BMJ Open 11 (1) (2021) e038137, https://doi.org/10.1136/bmjopen-2020-038137.
[49] R. Archer, et al., Assessing prognosis and prediction of treatment response in early rheumatoid arthritis: systematic reviews, Health Technol. Assess. 22 (66) (2018) 1–294, https://doi.org/10.3310/hta22660.
[50] R.C. Gao, et al., Association between sleep traits and rheumatoid arthritis: a mendelian randomization study, Front. Public Health 10 (2022) 940161, https://doi.org/10.3389/fpubh.2022.940161.
[51] A. Jiang, et al., A novel risk classifier to predict the in-hospital death risk of nosocomial infections in elderly cancer patients, Front. Cell. Infect. Microbiol. 13 (2023) 1179958, https://doi.org/10.3389/fcimb.2023.1179958.