Chinese primary knee osteoarthritis progression cohort (CPKOPC) to evaluate the progression of knee osteoarthritis in the Beijing population: a prospective cohort study protocol

Minwei Zhao, Yupeng Liang, Xinguang Wang, Lin Zeng, Hua Tian

ABSTRACT

Introduction Millions of patients are currently suffering from pain and dysfunction caused by osteoarthritis (OA), and billions of dollars have been invested into treatment. Because there is no effective treatment that can reverse the progression of knee OA, it is important to determine the risk factors that may influence the progression. However, although there are many studies that examine risk factors for progression, there are only a few that specifically focus on the impact of each risk factor for predicting progression of knee OA. This study aimed to develop a cohort of patients with primary knee OA in the Beijing area to establish models that identify the influence of each risk factor on the prediction of knee OA progression.

Methods and analysis This is a prospective, multicentre, hospital-based cohort study. The study population comprises 2000 patients with primary knee OA from the Beijing area. The recruitment and baseline visits started in December 2017 and will finish in November 2018. After baseline visits, the patients will be followed for 3 years or until the occurrence of primary outcomes. Demographic variables will be collected during the baseline visit. Influencing factors including occupational exposures, family history and treatment will be collected at baseline and each follow-up visit. The primary outcome measure is a comprehensive index which will be combined with clinical WOMAC score, imaging K-L grade and clinical outcomes. These data will also be collected at baseline and each follow-up visit.

Ethics and dissemination This study protocol has been approved by Peking University Third Hospital Medical Research Ethics Committee. All the eligible participants will give written informed consent. The findings will be published in peer-reviewed journals and presented at national or international conferences. Besides, the results will be disseminated to all participants via the social software 'WeChat'.

Trial registration number ChiCTR-ROC-17013790; prerestuls.

INTRODUCTION

Osteoarthritis (OA), also called degenerative osteoarthropathy, may severely influence quality of life. It can cause serious localised pain, deprive joint functions and eventually lead to disabilities. This leads to a huge medical and economic burden to society due to the increased proportion of the elderly population. Research shows there were 9.3 million patients suffering from symptomatic knee OA in the USA in 2005, and the global age-standardised prevalence of knee OA was 3.8% in 2010. In China, the overall prevalence of symptomatic knee OA was as high as 8.1% in 2012.

Currently, there is no effective treatment which can reverse progression of knee OA. Meanwhile, adverse reactions from symptomatic treatment, high costs and surgical risks from knee replacement make it hard to balance the costs and benefits of knee OA treatment. Under these circumstances, it would be ideal to restrain knee OA within the early stages to improve prognosis and reduce the burden on both medical and economic
resources. To accomplish this aim, analysis of the risk factors related to knee OA progression is necessary to make specific and targeted efforts at prevention.

The risk factors of OA have been studied extensively and various factors including age, genes, sex, hormones, biomechanics and local environment appear to have explicit relevance to the occurrence of OA. The progression of OA is also influenced by various factors. Generally, the systematic factors include age, race and heredity which influence susceptibility to OA and local factors including weight, biomechanics and local environment effect progression of OA. Thus, the synergistic effect of systematic and local factors influence OA progression.

However, there is still a lack of information about the impact of each risk factor for predicting progression of primary knee OA. To our knowledge, the first accepted model of knee OA progression risk factors was developed by Zhang et al in 2011, yet there has been no large-scale cohort studies or databases established in China according to a rapid review of Medline, Embase, Science Citation Index (SCI), Cumulative Index to Nursing and Allied Health Literature, Bioscience Information Service (BIOSIS), Cochrane database, China National Knowledge Infrastructure (CNKI) and Wanfang databases. In contrast, large-scale databases have already contributed greatly to preventive treatment for cancer and neurological, endocrine and cardiovascular disease. Moreover, assessment of prevention and treatment techniques using large-scale databases have a decided impact on health economic policy.

In conclusion, it is of great importance to establish a cohort of patients with primary knee OA to follow the progression of primary knee OA, analyse related risk factors and develop a predictive model for meaningful early prevention of primary knee OA progression. Therefore, the objectives of this study are to:

1. Develop a cohort of patients with knee OA to collect baseline and knee OA progression data.
2. Analyse progression risk factors associated with knee OA and develop a risk assessment model.
3. Verify risk prediction ability of the model using both internal and external validation.

METHODS AND ANALYSIS

Participants and outcomes

Study setting
This study is a multicentre, prospective cohort study of people with knee OA at early stages living in Beijing, sponsored by the Beijing Municipal Science and Technology Commission, located in Beijing, China. There are five clinical centres participating in the study including Peking University Third Hospital, Peking University First Hospital, Peking University People’s Hospital, Beijing Friendship Hospital and China–Japan Friendship Hospital.

Sample size estimate
The proportion of primary knee arthritis progression during the study was assumed to be 15% based on literature reports, and there are 16 predictive factors included in, which were confirmed by most of the studies. According to the multivariable logistic regression model sample size estimation method, 16×15=240 cases of knee OA progression are needed, so the total number of cases by final follow-up should be 240÷15%=1600 cases. Considering the possibilities of loss of follow-up, the sample size was expanded to 1.2 times. Therefore, the estimated sample size is 2000 cases.

Study population and eligibility
Recruitment of a baseline cohort of 2000 participants will be conducted between December 2017 and November 2018. The participants will be followed up at 12 months, 24 months, 30 months or 36 months, depending on the date of recruitment (table 1).

### Table 1 Items and procedures of the study at baseline and follow-up

<table>
<thead>
<tr>
<th></th>
<th>0 months</th>
<th>12 months</th>
<th>24 months</th>
<th>36 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informed consent</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inclusion/exclusion criteria</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demographic data</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk factors</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Intervention factors</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>WOMAC score</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>X-ray</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Functional MRI</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Biomarker</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Gait analysis data</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

’X’ indicates that the procedure is carried out.

WOMAC, Western Ontario and MacMaster Universities Osteoarthritis.
Routine and functional MRI scans will be performed on subjects with knee OA at the Peking University Third Hospital and China–Japan Friendship Hospital. The scanning device is a General Electric Company (GE, USA) MR scanner (field strength 3.0T). Routine scans include sagittal proton density-weighted image (PDWI), T1-weighted image, coronal PDWI and axial PDWI. Functional imaging includes sagittal T2 mapping and T1ρ. The specific method for obtaining knee cartilage measurements will be to use postprocessing functional imaging to obtain pseudocolour maps to identify the degree of cartilage damage in different areas, delineate the region of interest and measure the T2 and T1ρ values. The obtained knee cartilage measurements will be analysed after the subjects have completed all follow-up examinations.

In this study, CTX-II, which is the most concerning type of cartilage conversion biomarker, was selected for detecting progression of knee joint OA. To reduce the burden on participants and research convenience, urine samples will be used for CTX-II analysis by ELISA. Trained laboratory researchers will centrifuge the urine sample within 24 hours, take 1–2 mL of the supernatant, and deliver it to the orthopaedic clinical laboratory of Peking University First Hospital for ELISA kit analysis. If the sample cannot be delivered within 24 hours, it will be stored in a low temperature refrigerator until delivered.

The gait and related kinetic parameters of the subjects will be measured by using a three-dimensional (3D) gait analyser. The pace, step frequency and step size of the subjects will be observed, the duration of the lower limb support period and the percentage of support period will be recorded, and the knee adduction moment will be calculated to assess medial compartment load. Gait analysis will be carried out in the orthopaedic laboratory of Peking University Third Hospital.

### Eligibility criteria

Inclusion criteria was defined by: diagnosis of primary knee OA based on the 2007 revised of diagnostic criteria by the Chinese Medical Association Orthopaedic Branch (Table 2) and patients with a registered Beijing residence or permanent resident population in Beijing. Patients were excluded based on the following criteria: knee joint pain caused by diseases other than knee OA (infection, hip joint disease, lumbar disease and so on); secondary knee OA (trauma, rheumatoid arthritis, ankylosing spondylitis and so on); presence of severe internal medicine disease (American Society of Anesthesiologists’ classification score ≥3); inability to accomplish 3 years follow-up and refusal to sign the informed consent form.

### Measures

#### Exposure measures

The exposures of interest include demographic variables, risk factors and intervention factors. Demographic variables include age at OA diagnosis, sex, height, weight and ethnicity. Risk factors include occupational exposures (physical labour, knee-lifting and weight-bearing exercise, prolonged kneeling posture and so on), family history, history of knee trauma, lower limb muscle strength (femoral quadriceps, adductor muscle and hamstring muscle strength) and female menopause. Interventional factors include drug treatment (analgesic drugs, nutrient cartilage drugs, antiosteoporosis drugs and so on), physical therapy, Chinese medicine treatment, exercise therapy and surgical treatment.

#### Special measures

This study will also include three new techniques to predict the progression of knee OA: functional MRI, type II collagen C-terminal peptide (CTX-II) biomarker and gait analysis. These techniques are not widely used for this purpose, but research shows that they have potential to be important predictors.
imaging OA. Criteria to define patellofemoral imaging bone arthritis are as follows: grade 0 = no radiographic features of OA are present; grade 1 = doubtful joint space narrowing (JSN) and possible osteophytic lipping; grade 2 = definite osteophytes and possible JSN on anteroposterior weight-bearing radiograph; grade 3 = multiple osteophytes, definite JSN, sclerosis, possible bony deformity; grade 4 = large osteophytes, marked JSN, severe sclerosis and definite bony deformity.

The outcome indicator for this study is knee OA progression, which is defined as any of the following 3 at the 30-month follow-up point:
1. More than two on WOMAC score progress in one follow-up period (12 months).
2. K-L graded follow-up period ≥ 1 grade compared with baseline.
3. Surgical intervention (arthroscopy, osteotomy, single ankle replacement, patellofemoral joint replacement or total knee arthroplasty).

Study timeline
The study protocol consists of three main steps: screening and enrolment, baseline visit and follow-up visits.

Screening and enrolment
The screening of potential participants will be conducted by an orthopaedic surgeon at each clinical centre to assess the eligibility for the study according to the inclusion and exclusion criteria. When eligible, subjects will be informed to participate during a meeting in the hospital. During the meeting, they will receive the study protocol which also contains answers to common questions, and they will also be verbally informed by trained research personnel regarding the nature and purpose of the study, their rights to quit the study at any time and given time to decide whether or not to participate. After that, study participants will be required to sign a written informed consent.

Baseline visit
Once the written informed consent is obtained, the following information will be obtained:
1. Demographic information.
2. Risk factors.
3. Intervention factors.
4. WOMAC score.
5. X-ray (anteroposterior and lateral film).
6. Functional MRI.
7. Biomarker (CTX-II).
8. Gait analysis data.

Follow-Up
The following information will be collected at 12, 24, 30 and 36 months after enrolment:
1. Risk factors.
2. Intervention factors.
3. WOMAC score.
4. X-ray (anteroposterior and lateral film).
5. Functional MRI.

Data collection, management and analysis

Data collection methods
The study will develop a relevant case report form (CRF) based on the research plan. Trained researchers will load the data into the case report form in a timely, complete and accurate manner based on the original observation records of the research subjects. Then, the questionnaire will be reviewed and signed by an audit team and sent to the clinical research data administrator. Data entry will be done by two administrators using two computers independently, and the data will be compared twice. Problems found during the entry or comparison will be relayed to the inspector in a timely manner, and the researcher will be asked to explain. Questions and answers will be noted on a question form and kept for future reference.

Data verification and management
After all CRF are entered and verified, the data manager will run a database check report, which includes the completion of the study (including the list of shedding), the inclusion/exclusion criteria, the integrity inspection, the logical consistency inspection, outlier data inspection, time window inspection, intervention inspection and adverse event inspection.

At the review meeting, the main investigator, the representative of the collaboration unit, the inspector, the data administrator and the statistician will decide on the issues raised in the informed consent form and the database inspection report to write the audit report. The database will be locked at the same time.

Data archiving
After data entry and verification of the CRF are completed as required, it will be archived in the order of number for verification. Electronic data files include databases, inspection programme, analysis programme, analysis results, codebooks and documentation. Data will be stored in separate categories, and multiple backups will be stored on different disks or recording media for safekeeping in case of loss or damage. All the original files will be kept for a period of time according to corresponding regulations.

Quality control
Standardised training was conducted for all participating surgeons and radiologist regarding protocol for the CRF, physical examination, imaging examination and all other data collection and storage. In addition, all data were checked by research staff in our team to make sure all data are reliable.

Statistical methods
The demographic characteristics, general conditions, risk factors and intervention factors will be described by using the general statistical description method according to the distribution of the data. Nested case-control analysis
will be performed, and the subjects will be divided into knee OA progression group and non-progression group based on WOMAC score and imaging K-L grade. The differences of the demographic characteristics risk factors and intervention factors between the two groups will be compared. Several statistical methods will be used to establish the risk prediction model of knee OA progression, including logistic regression models, discriminant analysis and classification decision trees. K-fold cross validation will be performed to select the best prediction model based on the C-index of these prediction models. The external validation will use the pre-established knee osteoarthritis cohort data to determine the accuracy of the established model.

**PATIENT AND PUBLIC INVOLVEMENT**

Although patients were not explicitly involved in the development of the research questions, design, recruitment or conduct of this study, the research question and outcome measures are based on many years of clinical experience between all the authors. The participants will be encouraged to give feedback about their experience to their clinicians, who are part of the research team.

**DISCUSSION**

The study is going to establish a study cohort of knee OA, follow-up the progression of knee OA and analyse the relevant risk factors so as to establish a predictive model for knee OA progression and outcome, which will be of great value for the early prevention and treatment of primary knee OA. To date, the research on the progression of primary knee OA has revealed the influencing factors, but the importance of each influencing factor is still unclear. We still cannot evaluate the progression risk based on the patient's lifestyle and living environment. Furthermore, the risk of progression cannot be targeted to preventive interventions. By constructing a progression model, we can calculate the risk of OA progression for each patient, evaluate the contribution of each factor, perform targeted interventions for high-risk patients, such as appropriate weight loss treatments, exercise therapy and Chinese medicine treatment.

At present, there is a lack of mature and generally acknowledged risk prediction model for primary knee OA. The arthritis progression model study proposed by Zhang et al. in 2011 was based on 99 cases and 179 control groups. However, the scale of the study was relatively small, and the model's ability to go through external validation was poor (area under the receiver operating characteristic (ROC) curve=0.52). Additionally, the model failed to assess the risk of biomarker or cartilage function and to predict the progression of knee OA on a mechanism level.

To improve the accuracy of the prediction model, we plan to use the classification decision tree, discriminant analysis and other methods to establish a prediction model and use the K-fold cross-validation method to select the best prediction model. The original sample will be randomly divided into k equal-sized subsamples. Of the k subsamples, a single subsample will be retained as the validation data for testing the model, and the remaining k-1 subsamples will be used as training data. The cross-validation process will then be repeated k times, with each of the k subsamples used exactly once as the validation data by comparing the C-index between different prediction models. This will enable to select a more accurate prediction model. This method of cross-validation is relatively mature and efficient and will help to guide the selection of the optimal prediction model in this study. In addition, we will use an external cohort to validate our model. The database of 1294 patients with knee OA is established by the Peking University People's Hospital in Zhaoyunying Town Health Centre in Beijing.

To improve the sensitivity of the study, we will perform the following research design. We will use functional MRI, biological marker evaluation and gait analysis as secondary outcome indicators. Functional MRI will serve as a model observation index with good sensitivity and specificity, because it allows direct observation and evaluation of the cartilage, the most fundamental pathological structure of OA; the detection of related molecular biological markers (CTX-II, urinary-CTX-II) and 3D gait analysis will help observe the progression of OA from the viewpoint of molecular biology and biomechanics. These have the theoretical advantage of early and accurate assessment of the development of OA are of great significance for the further exploration of relevant mechanisms. This study also intends to combine the WOMAC score, KL grade imaging and clinical outcome assessment to evaluate knee OA progression. This comprehensive outcome index is expected to improve the sensitivity of the model within a limited follow-up time.

The study has the following limitations: (1) Short study time: the duration of this study is limited at 3 years because of the project requirements; hence, it might be difficult to observe the actual progression of OA within this limited time period. (2) Insufficient sample size: the sample size of this study design is only 2000, which is relatively small for establishing a progression model. In response to these two limitations, we will use unique statistical methods and outcome indicators to improve the accuracy and sensitivity of the research. Besides, we plan to continue to enrol new patients after the 3-year period to expand the database and improve the progression model. (3) The subjects enrolled in this study are permanent residents of Beijing, excluding the rural population. To improve the extrapolation of the model results, we will use external data including 1294 rural populations for external verification and establish models for both urban and rural populations or perform classification modelling according to the results.

**Contributors** MZ and YL are joint first authors. HT and MZ obtained funding. HT and MZ and LZ designed the study. YL, MZ, XW and LZ were involved in data...
collecting, follow-up and data analysing. YL and MZ drafted the manuscript. XW, LZ and HT contributed to critical revision of the manuscript for important intellectual content and approved the final version of the manuscript. All authors have read and approved the final manuscript. HT is the study guarantor.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Ethics approval** Peking University Third Hospital Medical Science Research Ethics Committee has approved (IRB00006761-M2017127) the protocol [version 1.0 (2017-5-20)].

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

**REFERENCES**


