

Prognostic Models for Knee Osteoarthritis: A Protocol for Systematic Review, Critical Appraisal and Meta-Analysis

Jingyu Zhong

Tongren Hospital Shanghai Jiaotong University School of Medicine <https://orcid.org/0000-0002-9817-2294>

Liping Si

Tongren Hospital Shanghai Jiaotong University School of Medicine

Guangcheng Zhang

Shanghai Sixth Peoples Hospital

Jiayu Huo

Shanghai Jiao Tong University

Yue Xing

Tongren Hospital Shanghai Jiaotong University School of Medicine

Yangfan Hu

Shanghai Sixth Peoples Hospital

Huan Zhang

Shanghai Jiao Tong University Medical School Affiliated Ruijin Hospital

Weiwu Yao (✉ yaoweiwuhuan@163.com)

Tongren Hospital Shanghai Jiaotong University School of Medicine

Protocol

Keywords: Knee, Osteoarthritis, Total knee arthroplasty, Prediction model, Prognosis, Systematic review, Meta-analysis

Posted Date: September 3rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-70145/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 May 19th, 2021. See the published version at <https://doi.org/10.1186/s13643-021-01683-9>.

Abstract

Background: Osteoarthritis is the most common degenerative joint disease diagnosed in clinical practice. It is associated with significant socioeconomic burden and poor quality of life, a large proportion of which is due to knee osteoarthritis (KOA), mainly driven by total knee arthroplasty (TKA). As the difficulty of being detected early and deficiency of disease-modifying drug, the focus of KOA is shifting to disease prevention and the treatment to delay its rapid progression. Thus, the prognostic prediction models are called for, to stratify individuals to guide clinical decision making. The aim of our review is to identify and characterize reported multivariable prognostic models for KOA which concern about three clinical concerns: (1) the risk of developing KOA in general population; (2) the risk of receiving TKA in KOA patients; and (3) the outcome of TKA in KOA patients who plan to receive TKA.

Methods: Studies will be identified by searching seven electronic databases. Title and abstract screening and full-text review will be accomplished by two independent reviewers. Data extraction instrument and critical appraisal instrument will be developed before formal assessment, and will be modified during a training phase in advance. Study reporting transparency, methodological quality, and risk of bias will be assessed according to Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement, CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) and Prediction model Risk Of Bias ASsessment Tool (PROBAST). Prognostic prediction models will be summarized qualitatively. Quantitative metrics on predictive performance of these models will be synthesized with meta-analyses if appropriate.

Discussion: Our systematic review will collate evidence from prognostic prediction models that can be used through the whole process of KOA. The review may identify models which are capable of allowing personalized preventative and therapeutic interventions to be precisely targeted at those individuals who are at the highest risk. To accomplish the prediction models to cross the translational gaps between an exploratory research method and a valued addition to precision medicine workflows, research recommendations relating to model development, validation or impact assessment will be made.

Systematic review registration: PROSPERO (registered, waiting for assessment, ID 203543)

Background

Osteoarthritis, a major source of pain, disability, and socioeconomic cost worldwide, is the most common degenerative joint disease leading to substantial and growing burden, and a large proportion of patients suffering from osteoarthritis is due to knee osteoarthritis (KOA) [1–3]. It has been estimated that healthcare costs of osteoarthritis accounts for about 1–2.5% of national gross domestic product, mainly driven by knee joint replacement, in particular, total knee arthroplasty (TKA) [2]. As the difficulty of being detected early and deficiency of disease-modifying drug [2, 3], the focus of KOA is shifting to disease prevention and the treatment to delay its rapid progression. Here, the prognostic prediction models are called for to distinguish individuals who are at higher risk of development or progression of KOA and who

are more likely to acquire better quality of life after TKA, which in turn could be used to guide clinical decision making.

Firstly, prevention is the best cure. Although the etiology of KOA has not been fully elucidated, a combination of risk factors is deemed to be related to this disease [4, 5], which allows the establishment of KOA risk prediction models. The evolving understanding of pathophysiological aspect of KOA [2, 3] is paralleled by improvements in prediction models, from only considering limited factors to a model combined clinical, genetic, biochemical and imaging information [6–10]. Prognostic models that showed moderate performance in evaluating KOA risk in general population may serve as a potential applicable tool for clinicians to stratify individuals by their risk level to provide suitable prevention strategy.

Secondly, current widely available diagnostic modalities do not fulfill the needs of clinicians to reduce the prognosis of KOA patients [11]. Therefore, the development and validation of prediction models that capable of identifying KOA patients at high risk of rapid progression is now recognized as a priority [12, 13]. TKA is the only available treatment option for KOA patients at the end stage and most of healthcare costs attributed to KOA is brought by this approach [2, 4]. Thus, it is preferable for KOA patients to delay TKA and to prolong the good health of their knees. Several models concerning TKA risk in KOA patients that conducted based on clinical information are reported, whose performance could be improved with introduction of imaging data [14–18]. Such models are necessitated for clinicians to pursue appropriate treatment options.

Thirdly, although TKA is a cost-effective surgical procedure that can advance quality of life [19], up to one third of KOA patients did not satisfy with their clinical outcomes [20]. A series of systematic reviews are conducted to report pooled survival of knee replacement [21], but a study that concentrate on prediction models for TKA outcomes has not been performed so far. The measures of prognostic models regarding TKA outcome vary from patient satisfaction to knee pain, stiffness, and dysfunction [22–25]. As the lack of agreement respecting indications for TKA currently [26], it is of significance to develop prediction models for aiding clinicians to KOA patient selection and therapeutic decision-making in TKA.

In spite of the huge amount of prediction models for KOA established, none of them is widely-accepted as an addition to precision medicine workflows. A systematic review of various models is helpful for improving their methodological quality and is necessary before their translation into clinical practice [27, 28]. It is timely to conduct a critical appraisal thorough specialized tools for prediction models for KOA [29–32]. Further, to provide a whole view of current prognostic models for KOA, we will include three sorts of models which run thorough clinical practice procedures of KOA [2, 3, 11, 27, 29].

In this study, we aim to identify reported multivariable prognostic models for KOA, which concern about three clinical questions: (1) to predict the risk of developing KOA in general population; (2) to predict the risk of receiving TKA in KOA patients; and (3) to predict the outcome of TKA in KOA patients who plan to receive TKA. We aim to map their characteristics, to critically appraise their reporting transparency, methodological quality, and risk of bias, and to meta-analyze their performance measures if possible.

Methods/design

Study design

This protocol is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols (PRISMA-P) statement [33] and the corresponding checklist can be found in the additional file 1. This protocol was registered on the International Prospective Register of Systematic Reviews (PROSPERO) (registered, waiting for assessment, ID 203543; additional file 2) [34].

This study will systematically review the prognostic models for development and prognosis of KOA. The framing of the review question, study identification, data collection, critical appraisal, data synthesis and results interpretation and reporting will be conducted according to previous guidelines and several developments in prediction model research methodology [29–32, 35–40] (Table 1).

Table 1
Timeline of the study

Stage of the review	Started	Completed	Rescores
Protocol drafting	Yes	No	PROSPERO, PRISMA-P, CHARMS, PICOTS, PROGRESS-3
Preliminary searches	Yes	No	PICOTS
Piloting of the study selection process	Yes	No	CHARMS, TRIPOD
Developing of review tools	Yes	No	Data extraction tool, Critical appraisal tool
Formal searches	No	No	PRESS, De-duplication guideline
Formal screening of search results against eligibility criteria	No	No	CHARMS, PICOTS, TRIPOD
Data extraction	No	No	Modified data extraction tool
Critical appraisal	No	No	Modified critical appraisal tool
Data analysis	No	No	Cochrane Library
Reporting	No	No	GRADE, PRISMA
<p>Note: CHARMS: CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies; GRADE: Grades of Recommendation, Assessment, Development, and Evaluation; PICOTS: Population, Intervention, Comparison, Outcome, Timing, Setting; PRESS: Peer Review of Electronic Search Strategies; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses; PRISMA-P: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols; PROBAST: Prediction model Risk Of Bias ASsessment Tool; PROGRESS: Prognosis Research Strategy; PROSPERO: International Prospective Register of Systematic Reviews; TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis.</p>			

Key items of this review are clarified with assistance of CHARMS checklist [31] (Table 2). A prognostic model will be defined as a combination of two or more predictors within statistical methods, machine learning methods or deep learning methods [35], which is used to predict the risk of the future outcome, and may help the health professionals and patients approach appropriate therapeutic decision. Studies investigated the association between a single risk factor and the outcome will be excluded, as they are limited in their utility for individual risk prediction. Specially, machine learning models in medical imaging, although they are usually based only on one modality, will be considered as multivariable, if multiple features have been extracted or deep learning methods have been employed. Studies reporting following types of prognostic models will be eligible for inclusion for our review: prediction models development with validation or external model validation. Studies that have developed prognostic models without validation will not be included into analysis, but records of these studies will be kept. We plan to systematically review prognostic models aiming (1) to predict KOA risk in general population; (2) to predict TKA risk in KOA patient; and (3) to predict TKA-related outcomes or complications in KOA patients intend to receive TKA, respectively; while studies reporting prognostic models with other objectives will not be considered.

Table 2
Framing of this systematic review using CHARMS key items

Key item	Model aim 1	Model aim 2	Model aim 3
1. Prognostic versus diagnostic prediction model	Future events: prognostic prediction models		
2. Intended scope of the review	Models to inform physicians' therapeutic decision making		
3. Type of prediction modelling studies	All study types: (1) prediction model development studies with internal or external validation; (2) external model validation studies with or without model updating.		
4. Target population to whom the prediction model applies	General population without KOA, with or without risk factors for KOA	KOA patient who has not receive TKA	KOA patient who plan to receive TKA
5. Outcome to be predicted	KOA risk	TKA risk	TKA outcomes
6. Time span of prediction	After the predictors collected, before the diagnosis of KOA	After the diagnosis of KOA, before the treatment of TKA	After the TKA
7. Intended moment of using the model	To predict the risk of KOA in general population	To predict the risk of receiving TKA after the diagnosis of KOA	To predict the TKA outcomes before TKA
Note: KOA, osteoarthritis; TKA, total knee arthroplasty			

Study inclusion

-Eligibility criteria

PICOTS (Population, Intervention, Comparison, Outcome, Timing, Setting) system will be used to frame the eligibility criteria and to guide selection of models with three different aims, separately [29, 30] (Table 3). PICOTS system is modified from PICO (Population, Intervention, Comparison, Outcome) system, additionally considers timing, i.e. specifically for prognostic models, when and over what time period the outcome is predicted; and setting, i.e. the intended role or setting of the prediction model.

Table 3
Eligibility criteria framed using the PICOTS system

PICTS system	Inclusion	Exclusion	Consideration
Model aim 1: To predict KOA risk in general population.			
Population	General population without KOA, with or without risk factors, asymptomatic or symptomatic	Population with KOA diagnosed by any criteria, patients with other knee diseases if they are not a predictor defined by study author.	General population from community, outpatient department, or pre-collected dataset will be considered for inclusion. Some studies in population with symptoms, such as knee pain, will be considered for inclusion if a diagnosis of KOA have not been established. Patients with other knee diseases will be excluded unless the condition is defined by study authors as a predictor for future KOA risk.
Index	Development and/or validation of a prognostic model for population without KOA to predict KOA risk	Diagnostic models for KOA	Prognostic model development with or without validation, and validation with or without updating will be considered for inclusion, if they are intended to predict KOA risk for general population. Diagnostic models will be excluded as our concern is to prevent KOA.
Comparator	Not applicable	Not applicable	As far as we know, a widely-adapted model for predicting KOA risk has not been established yet. Therefore, a comparison seemed to be impossible.
Outcomes	Future KOA diagnosis, KOA risk within a time period defined by the study's authors	Current KOA status	Most of current studies defined Kellgren and Lawrence grade ≥ 2 as KOA, while other studies may identify KOA patients with diagnostic codes. The effect measures for KOA will be as defined by the study's authors, their reference standard will be recorded.
Timing	KOA occurring after the predictors collected	Undiagnosed KOA before or at the moment the predictors collected	Included studies need to report on prediction models for future KOA occurring after the predictors collected. Prediction models for occurred undiagnosed KOA will be excluded.
Setting	Prognostic models that are intended to be used by healthcare professionals, in any clinic setting, at any time before the KOA diagnosis established.	Prognostic models that are intended to be used after or at the moment of establishing a diagnosis of KOA	Prognostic models that are intended to inform clinicians' therapeutic decision-making, i.e. prevention of KOA in high risk patients, will be included, to improve patient care. Prognostic models predicting progression of KOA patients will be excluded for this sub-question.

Note: KOA, osteoarthritis; TKA, total knee arthroplasty

PICTS system	Inclusion	Exclusion	Consideration
Model aim 2: To predict future TKA in KOA patient			
Population	KOA patient who has not receive TKA	KOA patient who has received TKA, undiagnosed KOA patients, general population without KOA, patients with other knee diseases	KOA patient diagnosed by any criteria, receiving any therapy except for TKA will be considered for inclusion. Patients with other knee diseases or without an established KOA diagnosis will be excluded.
Index	Development and/or validation of a prognostic model for KOA patient who has not receive TKA to predict necessity of TKA	Prognostic model for patient with other knee diseases, or to predict necessity of other therapeutic options, or symptoms.	Prognostic model development with or without validation, and validation with or without updating will be considered for inclusion, if they are intended to predict necessity of TKA for KOA patients. Prognostic models for patients with other knee diseases, or to predict necessity of other therapeutic options, or symptoms, will be excluded.
Comparator	Not applicable	Not applicable	As far as we know, a widely-adapted model for predicting future TKA in KOA patients has not been established yet. Therefore, a comparison seemed to be impossible.
Outcomes	Future TKA due to KOA, TKA risk within a time period defined by the study's authors	Necessity of other therapeutic options, or symptoms; TKA due to other knee diseases	As healthcare costs attributed to OA are driven largely by TKA, prognostic models of identifying patients with OA at high risk of future progression may be most useful for care healthcare professionals.
Timing	TKA after or at the moment of the diagnosis of KOA	TKA before the diagnosis of KOA, TKA in KOA patients who has received TKA	Included studies need to report on prediction models for future TKA after the diagnosis of KOA. Prediction models for general population will be excluded as they are less useful in practice. Prediction models for revision of TKA will also be excluded as our concern is to delay TKA.

Note: KOA, osteoarthritis; TKA, total knee arthroplasty

PICTS system	Inclusion	Exclusion	Consideration
Setting	Prognostic models that are intended to be used by healthcare professionals, in any clinic setting, at any time after the KOA diagnosis have been established, but before TKA	Prognostic models that are intended to be used before a diagnosis of KOA have been established	Prognostic models that are intended to inform clinicians' therapeutic decision-making, i.e. management of KOA to delay TKA.
Model aim 3: To predict TKA-related outcomes or complications in KOA patients intend to receive TKA			
Population	KOA patient who plan to receive TKA	KOA patient who has received TKA, undiagnosed KOA patients, general population without KOA, patients plan to receive TKA due to other knee diseases	KOA patient diagnosed by any criteria, planning to receive TKA will be considered for inclusion. Patients plan to receive TKA due to other knee diseases or without an established KOA diagnosis will be excluded.
Index	Development and/or validation of a prognostic model for KOA patient who plan to receive TKA to predict TKA-related outcomes or complications	Prognostic model for patient with other knee diseases who plan to receive TKA to predict outcomes or complications unrelated to TKA	Prognostic model development with or without validation, and validation with or without updating will be considered for inclusion, if they are intended to predict TKA-related outcomes or complications in KOA patients.
Comparator	Not applicable	Not applicable	As far as we know, a widely-adapted model for predicting TKA-related outcomes or complications in KOA patients plan to receive TKA has not been established yet. Therefore, a comparison seemed to be impossible.
Outcomes	TKA-related outcomes or complications	Outcomes or complications unrelated to TKA	As our aim is to select KOA patients suitable for TKA, only outcomes or complications related to TKA are useful for care healthcare professionals.

Note: KOA, osteoarthritis; TKA, total knee arthroplasty

PICTS system	Inclusion	Exclusion	Consideration
Timing	TKA-related outcomes or complications after the TKA	TKA-related outcomes or complications before the TKA	Psychological problems such as anxiety may occur before TKA; however, they are more likely to be recognized as predictors for poor outcomes or complications related to TKA.
Setting	Prognostic models that are intended to be used by healthcare professionals, in orthopedics setting, before TKA,	Prognostic models that are intended to be used after or at the moment of the TKA	Prognostic models that are intended to inform clinicians' and patients' therapeutic decision-making, i.e. to select KOA patients suitable for TKA, to prevent poor outcomes or complications in high risk patients.
Note: KOA, osteoarthritis; TKA, total knee arthroplasty			

We further established eligibility criteria concerning aspects other than the PICOTS system. (1) Study design: any study design including prospective or retrospective, randomized-controlled trial, observational study or case-control study, are acceptable. (2) Countries and regions: we will consider studies from all countries and regions. (3) Journal: we will consider studies from peer-reviewed journals of all research fields, which are representative of the high-quality studies on prognostic models for KOA. (4) Publish period: we will include only studies published after 2000, to display the current status of prediction modeling studies for KOA. Furthermore, the prediction model building approaches have significantly improved in the last two decades, particularly the machine learning methods and leading-edge deep learning methods. (5) Language: we will include studies published in English, Chinese, Japanese, German or French. One reviewer has expertise in those five languages. (6) Publication type: we will include only peer-reviewed full-text studies with original results, as they are expected to exhibit high-quality models and detailed methodology. Therefore, we will not consider abstracts only, conference abstracts, short communications, correspondences, letters or comments, and do not intend to search the grey literature. Any identified and relevant review articles will be used to identify eligible primary studies.

-Search strategy

We will search the following seven electronic databases from inception to 31 December 2020, including PubMed, Embase, the Cochrane Library, Web of Science, Scopus, SportDiscus, and Cumulative Index of Nursing and Allied Health Literature (CINAHL) [41-47]. SportDiscus is the leading bibliographic database for sports and sports medicine research; and CINAHL is the largest collection of full-text for nursing and allied health journals in the world. They will be included into electronic databases search because nursing and sports medicine professionals are also interested in the management of KOA patients; and these two databases were searched as routine in previous studies [48].

Search keywords will be selected from the MeSH terms and appropriate synonyms, based on the review question clarified by PICOTS system, including three concept terms: “knee”, “osteoarthritis” and “prediction model”. Each concept will be searched by MeSH term and free words combined with the OR Boolean operator, and then the three concepts will be combined with the AND Boolean operator. For each database, keywords will be translated into controlled vocabulary (MeSH, Emtree, and others), and will be chosen from free text. We will take search strategies in former studies as reference [48] and will co-design the search strategy. The search strategies will be tested for eligibility by two reviewers before formal search. The sample for the search strategy will be presented in additional file 3.

The formal search will be performed by two same reviewers according to the PRESS guideline [36]. In case of uncertainties, a third reviewer was consulted to reach final consensus. The reference list of included studies and relevant reviews will be hand-searched for additional potentially relevant citations. However, we do not intend to search grey literature due to concerns on their methodological quality.

-Data management

We will use Endnote reference manager software version X9.2 (Clarivate Analytics, Philadelphia, PA, USA) [49] to merge the retrieved studies. Duplicates will be removed using a systematic, rigorous and reproducible method utilizing a sequential combination of fields including author, year, title, journal and pages [37]. We will use a free online Tencent Document software (Tencent, Shenzhen, China) [50] to manage records through-out the review, to make sure all reviewers follow the latest status of the review process timely and to ensure two senior reviewers can supervise the process remotely during the difficult period of coronavirus disease 2019 pandemic.

-Study selection

Two independent reviewers will screen the titles and abstracts of all the potential records to identify all relevant studies using the pre-defined inclusion and exclusion criteria. In case of an unavailable abstract, full-text articles will be obtained unless the title is clearly irrelevant. Two same reviewers will obtain the full-text and supplementary materials of all selected records, and will thoroughly read them independently, to further determine their eligibility before extracting data. The corresponding authors of potential records may be contacted to request the full-text if it is not available otherwise. Disagreements will be resolved by consensus to reach the final decision, with assistance from our review group consist of a computer engineer with experience in prediction model building, an orthopedist with experience in OA management, and musculoskeletal radiologists.

Data collection

-Data extraction

We will develop a data extraction instrument for study data based on several previous systematic reviews of prediction model [51-53]. As the reviewers have different levels of experience and knowledge, the items

listed will be reviewed and discussed to ensure that all reviewers had clear knowledge of the procedures. A training phase will be introduced before the formal extraction.

During the training phase, two randomly chosen articles from all articles fulfilled the inclusion criteria for discussion will be used to train two independent reviewers. They will thoroughly read the two randomly chosen articles including the supplementary materials, and will measure each study independently. A structured data collection instrument will be modified and used to help them reach agreement. Disagreements will be discussed in order to achieve a shared understanding of each parameter. This pre-defined and piloted data extraction instrument will be used in the formal data extraction phase.

During the formal extraction phase, two independent reviewers will thoroughly read all articles including the supplementary materials, to extract the data from the studies to describe their characteristics. Any disagreement will be resolved by discussion to reach a consensus and consultation with other members of our review group if required. Missing data will be obtained from the authors wherever possible; studies with insufficient information will be noted.

-Critical appraisal

We will develop a critical appraisal instrument according to TRIPOD statement, CHARMS checklist and PROBAST tool [30-32]. The TRIPOD is a set of recommendations, deemed essential for transparent reporting of a prediction model study, and allows the quality evaluation and potential usefulness analysis. The CHARMS checklist identifies eleven domains to facilitate a structured critical appraisal of primary studies on prediction models, mainly focus on the methodological quality of included models. The PROBAST tool is designed for assessing the risk of bias and applicability concerning four domains, i.e. participants, predictors, outcome, and analysis, with a total of 20 signaling questions. These three instruments, although focus on different aspects of prediction model studies, overlap each other in several domain and items. Therefore, we will merge them into a critical appraisal instrument to reduce the workload during the systemic critical evaluation.

During the development period of this instrument, we also considered machine learning and deep learning relevant checklists, e.g. radiomics quality score [54], Checklist for Artificial Intelligence in Medical Imaging [55], Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research [56], all of which are specialized assessment tools for cutting-edge artificial intelligence models. However, they include many items that may not be available for prediction models built with traditional statistical methods based on clinical characteristics, laboratory examinations, or genetic factors. On the other hand, TRIPOD, CHARMS and PROBAST have been already proved suitable for assessing prediction models using artificial intelligence methods [53]. Thus, we will choose three more widely-adapted and more extensively-accepted tools, to develop our critical appraisal instrument.

A similar training phase is introduced before the formal critical appraisal, to ensure its eligibility and to achieve a shared understanding of each parameter. During the formal evaluation phase, two independent

reviewers will assess all the articles and corresponding supplementary materials, to measure and rate all studies according to established criteria. Any disagreement will be solved as described before.

-Data pre-processing

The necessary results or performance measures and their precision are needed to allow quantitative synthesis of predictive performance of the prediction model under study [29]. However, model performance measurements vary among reported prediction model studies, and sometimes are unreported or inconsistent for further analysis. In cases where pertinent information is not reported, efforts will be made to contact study authors to request this information. If there is any non-response, missing performance measures and their measures of precision will be calculated if possible according to the methods previously described [29]. If this is impossible due to limited data, the exclusion of the study will be determined by discussion among the reviewers.

Data synthesis

The data synthesis process will be guided by several methodological reference books and guidelines [57-61]. Two reviewers of this study have significant expertise in statistics and meta-analysis methods that would be used in this review. In case of doubt, the reviewers will discuss to approach consensus or consult a statistician for advice.

-Qualitative synthesis

All extracted data on prediction models will be narratively summarized and the key findings will be tabulated to facilitate comparison according to the PICOTS system [30], and in particular, what prediction factors were included in different models, when and how the included variables were coded, what the outcomes of models were, the reported predictive accuracy of the model and whether the model was validated internally and/or externally, and if so, how. Models relating to different aims will be considered separately. Two most common statistical measures of predictive performance, discrimination and calibration will be reported when published or approximated using published methods [30]. Other means such as sensitivity, specificity, positive predictive value and negative predictive value will also be included if reported [30]. Individual results of CHARMS, TRIPOD, and PROBAST and the overall reporting transparency, methodological quality, and risk of bias will be reported [30-32].

-Quantitative synthesis

The statistics analysis will be performed via SPSS software version 26.0 (SPSS Inc., Chicago, IL, USA) [62]. P-value < 0.05 will be recognized as statistical significance, unless otherwise specified. The elements of TRIPOD will be treated as binary categorical variables, with their inter-rater agreement assessed by the Cohen's kappa statistic [63]. The elements of CHARMS and PROBAST includes ordinal categories with more than two possible ratings, therefore, Fleiss' kappa statistic will be used to assess their inter-rater agreement [64]. The summed TRIPOD rating will be treated as a continuous variable, and their inter-rater agreement will be assessed using the interclass correlation coefficient (ICC) [65]. Further, we will provide

correlation information among these three instruments to present whether they are complimentary critiques [66], where possible.

-Meta-analysis

The meta-analysis will be conducted via Stata/SE software version 15.1 (Stata Corp., College Station, TX, USA) with the metan, midas and metandi packages [67-70] and any other packages depending on the data we extract. The plan of meta-analysis will be dependent on the studies identified in the systematic review. If a similar clinical question was assessed repeatedly in a large enough subset of the included studies, meta-analysis will be considered to jointly summarize calibration and discrimination statistics with their 95% confidence intervals to obtain average model performance. Relevant forest plots and a hierarchical summary receiver operating characteristic (HSROC) curve will be obtained to visually show the model performance [71].

-Heterogeneity assessment

For assessment of heterogeneity between the meta-analyzed studies, the Cochran's Q and the I^2 statistic will be calculated [72]. Difference between the 95% confidence region and prediction region in the HSROC curve was used to visually assess the heterogeneity, and a large difference indicate the presence of heterogeneity [71]. Potential sources of heterogeneity will be investigated by means of meta-regression or subgroup analysis if there are > 10 studies included in the meta-analysis [73].

-Publication bias assessment

Publication biases arise when the dissemination of research findings is influenced by the nature and direction of results. A Deeks funnel plot will be generated to visually assessed publication bias if there are > 10 studies included in the meta-analysis [74]. An Egger's test was performed to assess the publication bias and a p-value > 0.10 indicated a low publication bias [75]. A Deeks funnel plot asymmetry test was also constructed to explore the risk of publication bias, and a p-value > 0.10 indicated a low publication bias [76]. The trim and fill method will be conducted to estimate the number of missing studies [77].

-Subgroup analysis

We plan to carry out following subgroup analyses regardless of heterogeneity. (1) the type of model validation: internal validation or external validation; (3) the predictor of model: clinical characteristics, laboratory examinations, genetic factors, objective or quantitative-extracted imaging feature, or their combinations. (4) the method of prognostic model building: statistic method, machine learning method, or deep learning method, etc. These subgroups were selected to display the current strengths and limitation of studies in prediction model for KOA. Further subgroup analysis will depend on the data extracted.

-Sensitivity analysis

Sensitivity analyses will be performed by excluding studies with high risk of bias assessed by the PROBAST tool (at least 4/7 domain to be high), studies with high methodological quality assessed by the CHARMS checklist (at least 6/11 domain to be high), and studies with low reporting transparency assessed by the TRIPOD statement (at least half of available items not to be mentioned), to explore their influence on effect size. This analysis will be a narrative summary that covers the same elements as the primary analysis if appropriate.

Reporting and dissemination

The results of the review will be reported guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [38]. The confidence in estimates will be determined according to the GRADE approach (Grades of Recommendation, Assessment, Development, and Evaluation) [39,40]. The approval of ethics and consent to participate are not required for our study due to its nature of systematic review and meta-analysis. Our findings will be disseminated through peer-reviewed publications, and presentation at conferences if possible.

Discussion

This systematic review will identify all published prognostic prediction models for three important KOA-related clinical questions. These prognostic prediction models will be comprehensively summarized and critically appraised. Their performance will be meta-analyzed if appropriate, and further compared across pre-defined subgroups.

The major strength of our study is that it will provide a bird view of prognosis prediction model for this disease, and may point out future research directions for this field. Next, our review team is composed of musculoskeletal radiologists, an orthopedist and a computer engineer, which allows us to share our knowledge and expertise. We will introduce a training phase to reach a better understanding of included studies and used assessment tools. Then, the data extraction and critical appraisal instruction may become a reference for future reviews. Finally, we will calculate the inter-rater agreements which is seldom reported by previous reviews. This may improve the transparency and quality of our review.

A limitation of our study is that we are not able to search all the databases, restricting publication date to 2000–2020 and languages, as well as the grey literature. Besides, we will exclude several existing models for KOA patients regarding valuable aims apart from our set three clinical questions, such as knee pain, KOA progression, and response to other treatments. Moreover, we will only consider models predicting clinical outcomes, but neither socioeconomic burdens nor cost-effective aspects which are important in model practical translation. In addition, the instruments that we will use have limitations. While the sum score of TRIPOD is a quantitative metric, the CHARMS and PROBAST are qualitative scores and therefore less easily interpretable.

To summarize, our systematic review will be an important step towards developing and applying prognostic prediction models that can be used through the whole process of KOA. This will allow

personalized preventative and therapeutic interventions to be precisely targeted at individuals at highest risk, and to avoid harm and additional expense for those who is not.

Abbreviations

CHARMS: CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies;

CINAHL: Cumulative Index of Nursing and Allied Health Literature;

GRADE: Grades of Recommendation, Assessment, Development, and Evaluation;

ICC: interclass correlation coefficient;

KOA: knee osteoarthritis;

PICO: Population, Intervention, Comparison, Outcome;

PICOTS: Population, Intervention, Comparison, Outcome, Timing, Setting;

PRESS: Peer Review of Electronic Search Strategies;

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses;

PRISMA-P: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols;

PROBAST: Prediction model Risk Of Bias ASsessment Tool;

PROSPERO: International Prospective Register of Systematic Reviews;

TKA: total knee arthroplasty;

TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

Not applicable

Competing interests

The authors declare that they have no competing interests.

Funding

This work is supported by the National Natural Science Foundation of China (81771790) and the Medicine and Engineering Combination Project of Shanghai Jiao Tong University (YG2019ZDB09). They played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors' contributions

JZ, LS and GZ originally conceptualized the study. All the authors contributed to the development of protocol. JZ drafted the original manuscript. JH brought expertise in prediction model building. GZ brought expertise in clinical managements of osteoarthritis. JZ and LS did the preliminary searches, piloted the study selection process, and developed data analysis strategy, with assistance from YX, YH, JH and GZ. GZ polished English language of the manuscript. WY and HZ supervised the protocol developing process. WY acquired the funding and is the guarantor of this manuscript. All authors critically reviewed and approved the final version of manuscript.

Acknowledgements

The authors would like to thank editors and reviewers for their kindness and enlightening comments on this manuscript.

References

1. Safiri S, Kolahi AA, Smith E, et al. Global, regional and national burden of osteoarthritis 1990-2017: a systematic analysis of the Global Burden of Disease Study 2017. *Ann Rheum Dis.* 2020;79(6):819-828.
2. Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. *Lancet.* 2019;393(10182):1745-1759.
3. Martel-Pelletier J, Barr AJ, Cicuttini FM, et al. Osteoarthritis. *Nat Rev Dis Primers.* 2016;2:16072.
4. Palazzo C, Nguyen C, Lefevre-Colau MM, Rannou F, Poiraudeau S. Risk factors and burden of osteoarthritis. *Ann Phys Rehabil Med.* 2016;59(3):134-138.
5. Cooper C, Snow S, McAlindon TE, et al. Risk factors for the incidence and progression of radiographic knee osteoarthritis. *Arthritis Rheum.* 2000;43(5):995-1000.
6. Losina E, Klara K, Michl GL, Collins JE, Katz JN. Development and feasibility of a personalized, interactive risk calculator for knee osteoarthritis. *BMC Musculoskelet Disord.* 2015;16:312.

7. Kerkhof HJ, Bierma-Zeinstra SM, Arden NK, et al. Prediction model for knee osteoarthritis incidence, including clinical, genetic and biochemical risk factors. *Ann Rheum Dis*. 2014;73(12):2116-2121.
8. Zhang W, McWilliams DF, Ingham SL, et al. Nottingham knee osteoarthritis risk prediction models. *Ann Rheum Dis*. 2011;70(9):1599-1604.
9. Yoo TK, Kim DW, Choi SB, Oh E, Park JS. Simple Scoring System and Artificial Neural Network for Knee Osteoarthritis Risk Prediction: A Cross-Sectional Study. *PLoS One*. 2016;11(2):e0148724.
10. Joseph GB, McCulloch CE, Nevitt MC, et al. Tool for osteoarthritis risk prediction (TOARP) over 8 years using baseline clinical data, X-ray, and MRI: Data from the osteoarthritis initiative. *J Magn Reson Imaging*. 2018;47(6):1517-1526.
11. Jamshidi A, Pelletier JP, Martel-Pelletier J. Machine-learning-based patient-specific prediction models for knee osteoarthritis. *Nat Rev Rheumatol*. 2019;15(1):49-60.
12. Bruyère O, Cooper C, Arden N, et al. Can we identify patients with high risk of osteoarthritis progression who will respond to treatment? A focus on epidemiology and phenotype of osteoarthritis. *Drugs Aging*. 2015;32(3):179-187.
13. Arden N, Richette P, Cooper C, et al. Can We Identify Patients with High Risk of Osteoarthritis Progression Who Will Respond to Treatment? A Focus on Biomarkers and Frailty. *Drugs Aging*. 2015;32(7):525-535.
14. Chan WP, Hsu SM, Huang GS, Yao MS, Chang YC, Ho WP. Creation of a reflecting formula to determine a patient's indication for undergoing total knee arthroplasty. *J Orthop Sci*. 2010;15(1):44-50.
15. Yu D, Jordan KP, Snell KIE, et al. Development and validation of prediction models to estimate risk of primary total hip and knee replacements using data from the UK: two prospective open cohorts using the UK Clinical Practice Research Datalink. *Ann Rheum Dis*. 2019;78(1):91-99.
16. Tiulpin A, Klein S, Bierma-Zeinstra SMA, et al. Multimodal Machine Learning-based Knee Osteoarthritis Progression Prediction from Plain Radiographs and Clinical Data. *Sci Rep*. 2019;9(1):20038.
17. Tolpadi AA, Lee JJ, Pedoia V, Majumdar S. Deep Learning Predicts Total Knee Replacement from Magnetic Resonance Images. *Sci Rep*. 2020;10(1):6371.
18. Leung K, Zhang B, Tan J, et al. Prediction of Total Knee Replacement and Diagnosis of Osteoarthritis by Using Deep Learning on Knee Radiographs: Data from the Osteoarthritis Initiative. *Radiology*. 2020;192091.
19. Murray DW, MacLennan GS, Breeman S, et al. A randomised controlled trial of the clinical effectiveness and cost-effectiveness of different knee prostheses: the Knee Arthroplasty Trial (KAT). *Health Technol Assess*. 2014;18(19):i-viii.
20. Bullens PH, van Loon CJ, de Waal Malefijt MC, Laan RF, Veth RP. Patient satisfaction after total knee arthroplasty: a comparison between subjective and objective outcome assessments. *J Arthroplasty*. 2001;16(6):740-747.

21. Evans JT, Walker RW, Evans JP, Blom AW, Sayers A, Whitehouse MR. How long does a knee replacement last? A systematic review and meta-analysis of case series and national registry reports with more than 15 years of follow-up. *Lancet*. 2019;393(10172):655-663.
22. Lungu E, Desmeules F, Dionne CE, Belzile EL, Vendittoli PA. Prediction of poor outcomes six months following total knee arthroplasty in patients awaiting surgery. *BMC Musculoskelet Disord*. 2014;15:299.
23. Pua YH, Seah FJ, Clark RA, Poon CL, Tan JW, Chong HC. Development of a Prediction Model to Estimate the Risk of Walking Limitations in Patients with Total Knee Arthroplasty. *J Rheumatol*. 2016;43(2):419-426.
24. Van Onsem S, Van Der Straeten C, Arnout N, Deprez P, Van Damme G, Victor J. A New Prediction Model for Patient Satisfaction After Total Knee Arthroplasty. *J Arthroplasty*. 2016;31(12):2660-2667.
25. Shim J, Mclernon DJ, Hamilton D, Simpson HA, Beasley M, Macfarlane GJ. Development of a clinical risk score for pain and function following total knee arthroplasty: results from the TRIO study. *Rheumatol Adv Pract*. 2018;2(2):rky021.
26. Cross WW 3rd, Saleh KJ, Wilt TJ, Kane RL. Agreement about indications for total knee arthroplasty. *Clin Orthop Relat Res*. 2006;446:34-39.
27. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35(29):1925-1931.
28. Panken G, Verhagen AP, Terwee CB, Heymans MW. Clinical Prediction Models for Patients With Nontraumatic Knee Pain in Primary Care: A Systematic Review and Internal Validation Study. *J Orthop Sports Phys Ther*. 2017;47(8):518-529.
29. Debray TP, Damen JA, Snell KI, Ensor J, Hooft L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. *BMJ*. 2017;356:i6460.
30. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015;350:g7594.
31. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med*. 2014;11(10):e1001744.
32. Moons KM, Wolff RF, Riley RD, et al. Probast: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med*. 2019;170(1):W1–W33.
33. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4:1.
34. National Institute for Health Research. International Prospective Register of Systematic Reviews (PROSPERO). Available from: <https://www.crd.york.ac.uk/prospero>.
35. Steyerberg EW, Moons KG, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med*. 2013;10(2):e1001381.

36. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS peer review of electronic search strategies: 2015 guideline statement. *J Clin Epidemiol.* 2016;75:40–6.
37. Bramer WM, Giustini D, de Jonge GB, Holland L, Bekhuis T. De-duplication of database search results for systematic reviews in EndNote. *J Med Libr Assoc.* 2016;104(3):240–3.
38. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ.* 2009;339:b2535.
39. Iorio A, Spencer FA, Falavigna M, Alba C, Lang E, Burnand B, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. *BMJ.* 2015;350:h870.
40. Schünemann H, Brożek J, Guyatt G, Oxman A, Editors. GRADE handbook for grading quality of evidence and strength of recommendations. The GRADE Working Group; 2013. <http://training.cochrane.org/resource/grade-handbook>. Accessed 20 Jul 2020.
41. National Library of Medicine, National Center for Biotechnology Information. PubMed. Available from: <https://pubmed.ncbi.nlm.nih.gov>.
42. Embase. Available from: <https://www.embase.com>.
43. Cochrane Library. Cochrane Library. Available from: <https://www.cochranelibrary.com>.
44. Clarivate Analytics. Web of Science. Available from: <http://www.isiknowledge.com>.
45. Scopus. Available from: <https://www.scopus.com>.
46. EBSCO Information Services. SPORTDiscus. Available from: <http://search.ebscohost.com>.
47. EBSCO Information Services. Cumulative Index of Nursing and Allied Health Literature (CINAHL). Available from: <http://search.ebscohost.com>.
48. Hislop AC, Collins NJ, Tucker K, Deasy M, Semciw AI. Does adding hip exercises to quadriceps exercises result in superior outcomes in pain, function and quality of life for people with knee osteoarthritis? A systematic review and meta-analysis. *Br J Sports Med.* 2020;54(5):263-271.
49. Clarivate Analytics. EndNote version X9.2. 2018. Available from: <https://endnote.com>.
50. Tencent Document. 2018. Available from: <https://docs.qq.com>.
51. Bellou V, Belbasis L, Konstantinidis AK, Tzoulaki I, Evangelou E. Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal. *BMJ.* 2019;367:l5358.
52. Gerry S, Bonnici T, Birks J, et al. Early warning scores for detecting deterioration in adult hospital patients: systematic review and critical appraisal of methodology. *BMJ.* 2020;369:m1501.
53. Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ.* 2020;368:m689.
54. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* 2017;14(12):749-762.
55. Mongan J, Moy L, Kahn CE Jr. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell.* 2020;2(2):e200029.

56. Luo W, Phung D, Tran T, et al. Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View. *J Med Internet Res*. 2016;18(12):e323.
57. Cochrane methods prognosis. Tools. Cochrane methods prognosis; 2020. Available via <https://methods.cochrane.org/prognosis/tools>. Accessed 20 Jul 2020.
58. Cochrane methods screening and diagnostic tests. Handbook for DTA Reviews. Cochrane methods screening and diagnostic tests; 2020. Available via <https://methods.cochrane.org/sdt/handbook-dta-reviews>. Accessed 20 Jul 2020.
59. Zhang TS, Zhong WZ, Li B. Applied methodology for evidence-based medicine, 2nd edn. Changsha: Central South University Press; 2014.
60. Zhang TS, Dong SJ, Zhou ZR. Advanced meta-analysis in Stata. Shanghai: Fudan University Press; 2016.
61. Debray TP, Damen JA, Snell KI, Ensor J, Hooft L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. *BMJ*. 2017;356:i6460.
62. International Business Machines Corporation. SPSS Statistics version 26.0. 2019. Available from: <https://www.ibm.com/products/spss-statistics>.
63. Cohen JA. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20(1):37-46.
64. Marasini D, Quatto P, Ripamonti E. Assessing the inter-rater agreement for ordinal data through weighted indexes. *Stat Methods Med Res*. 2016;25:2611–2633.
65. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155–163.
66. Fornaçon-Wood I, Faivre-Finn C, O'Connor JPB, Price GJ. Radiomics as a personalized medicine tool in lung cancer: Separating the hope from the hype. *Lung Cancer*. 2020;146:197-208.
67. Stata Corporation. Stata/SE version 15.1. 2019. Available from: <https://www.stata.com>.
68. Harris R, Bradburn M, Deeks J, et al. METAN: Stata module for fixed and random effects meta-analysis. Statistical Software Components S456798, Boston College Department of Economics, revised 23 Sep 2010.
69. Dwamena B. MIDAS: Stata module for meta-analytical integration of diagnostic test accuracy studies. Statistical Software Components S456880, Boston College Department of Economics, revised 05 Feb 2009.
70. Harbord RM, Whiting P. Metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression. *Stata J* 9:211–29.
71. Macaskill P. Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis. *J Clin Epidemiol*. 2004;57(9):925–932.
72. Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index?[J]. *Psychol Methods*. 2006;11(2):193-206.
73. Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI. Anti-lymphocyte antibody induction therapy study group. Individual patient versus group-level data meta-regressions for the investigation

- of treatment effect modifiers: ecological bias rears its ugly head. *Stat Med.* 2002;21:371-87.
74. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol.* 2005;58:882–893.
75. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ.* 1997;315(7109):629-634.
76. Song F, Khan KS, Dinnes J, Sutton AJ. Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. *Int J Epidemiol.* 2002;31: 88–95
77. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics.* 2000;56(2):455-463.

Supplementary Files

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

- [PROSPEROregistrationID203543.pdf](#)
- [ModelinOASamplesearchstrategies.docx](#)
- [ModelinOAPRISMAPchecklist.docx](#)